Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# A Data-Driven Approach for Indirect Estimation of INLmax_abs Using Dynamic Parameters as Input Features

**Rongjie Li**
School of Automation Engineering, University of Electronic Science and Technology of China
**Houjun Wang**
Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
**Pu Pu**
School of Automation Engineering, University of Electronic Science and Technology of China
**Jia Li**
Shence chips testing and certification co.,LTD
**Guibing Zhu**
Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
**Wanyu Yang**
yangwanyu@uestc.edu.cn

Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China

**Research Article**

**Keywords:** INLmax_abs Estimation, Dynamic parameters, Hybrid KNN-SVR Model, Analog-to-Digital Converters

**Posted Date:** November 3rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10974048/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# A Data-Driven Approach for Indirect Estimation of INL*max_abs* Using Dynamic Parameters as Input Features

**Rongjie Li[a], Houjun Wang[b,d], Pu Pu[a], Jia Li[c], Guibing Zhu[b,d] and Wanyu Yang[b,d]**

**Corresponding author：Wanyu Yang**

**E-mail: yangwanyu@uestc.edu.cn**

[a]School of Automation Engineering, University of Electronic Science and Technology of China, 611731, Chengdu, Sichuan, China

[b]Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen, 518000, Guangdong, China

[c]Shence chips testing and certification co.,LTD., Shenzhen, 518000, Guangdong, China

[d]Shenzhen Gaoyan Integrated Circuit Testing Technology Co., Ltd., Shenzhen, 518048, Guangdong, China

## Abstract

INL testing based on the code-density histogram requires a large number of samples, which significantly degrades testing efficiency. To address this issue, this paper proposes a data-driven indirect estimation approach for INL*max_abs*. In particular, it takes dynamic parameters such as SNR, SINAD, THD, SFDR, and harmonics as input features, and constructs a machine learning model to estimate INL*max_abs*. For model construction, a Hybrid KNN-SVR model is developed by integrating the local learning capability of KNN with the global fitting advantage of SVR, thereby improving robustness across datasets with different distributions. The experimental results show that the RMSE values for INL*max_abs* estimation on the five ADC datasets are 0.0532, 0.0607, 0.0301, 0.2006, and 0.2083 LSB, respectively. Furthermore, the Hybrid KNN-SVR model demonstrates performance comparable to or even better than complex models such as XGBoost

and ANN. This work provides a valuable basis for the rapid assessment of INL performance.

**Keywords：** INL*max_abs* Estimation, Dynamic parameters, Hybrid KNN-SVR Model, Analog-to-Digital Converters

## Nomenclature

| | |
|---|---|
| INL*max_abs* | Maximum Absolute Value of the Integral Nonlinearity |
| SNR | Signal-to-Noise Ratio |
| SINAD | Signal-to-Noise and Distortion Ratio |
| THD | Total Harmonic Distortion |
| SFDR | Spurious-Free Dynamic Range |

## 1 Introduction

Accurate and rapid circuit testing is essential for ensuring product quality and improving test efficiency [1,2]. The nonlinear characteristics of analog-to-digital converters (ADCs) are primarily characterized by static and dynamic parameters. As one of the key static parameters, integral nonlinearity (INL) quantifies the deviation of the actual ADC transfer characteristic from the ideal transfer characteristic [3]. However, as ADC resolution and dynamic performance continue to improve, the time and equipment costs required for INL testing increase significantly [4,5,6]. Consequently, the optimization of INL testing has attracted considerable attention.

From the perspective of test equipment cost, INL testing requires the signal generator to provide a highly linear ramp signal or a high-purity sinusoidal signal to ensure measurement accuracy. When the performance of the signal generator cannot satisfy the test requirements, nonidealities in the excitation signal itself can degrade the accuracy of INL measurements. To address this issue, extensive research has focused on identifying and compensating for errors in the test stimulus, thereby reducing the dependence on high-performance signal sources [7,8].

From the perspective of test time, the number of samples required by the code density histogram increases exponentially as the resolution increases. To improve the efficiency of INL testing, existing strategies mainly include:1) Optimizing the test stimulus [9,10]; 2) Testing INL

using as few samples as possible [11-15]; 3) Indirectly fitting the INL curve through dynamic response [16-18]. In the early stages on static parameter testing using DC signals, a method was proposed to optimize the test stimulus by applying multiple DC components simultaneously, forming a staircase waveform for INL testing [9]. Furthermore, based on the sine histogram test (SHT), researchers have proposed the Fast Sine Wave Histogram Testing (FSHT) method [12] as well as a wavelet decomposition and reconstruction technique [13], both of which achieve INL curve testing while significantly reducing the number of required samples.

The above strategies require statistical analysis of the ADC output codes. In contrast, other studies exploit the relationship between INL and harmonic distortion to indirectly reconstruct the INL curve by extracting harmonic components from the output spectrum. The main implementations of this strategy include polynomial and discrete Fourier series. The polynomial is suitable for fitting smooth INL curves [16], whereas the discrete Fourier series can be employed to fit sharply varying INL curves [17,18]. However, when accurately fitting highly fluctuating INL curves, the number of Fourier coefficients increases rapidly, and higher order harmonic components are also required.

With the widespread application of machine learning techniques in integrated circuit testing, the indirect estimation of key device parameters using regression-based methods has attracted increasing attention [19-21]. Previous studies have shown that statistical correlations and analytical mapping relationships exist between the static and dynamic parameters of ADCs [22-24]. Meanwhile, from a physical perspective, poorer ADC dynamic performance is generally associated with more severe nonlinearities, which in turn lead to a larger variation range of the INL curve. Moreover, in practical ADC testing, the maximum absolute value of INL, denoted as INL*max_abs*, is commonly used as a specification metric for device evaluation. Therefore, this work proposes a machine learning based strategy for indirect estimation of INL*max_abs*. This strategy uses nonlinear parameters extracted from dynamic responses, such as SNR, SINAD, THD, SFDR, and harmonic components, as model input features to construct an INL*max_abs* estimation model. Thus, INL*max_abs* can be rapidly estimated solely from dynamic test results without applying a dedicated test stimulus during ADC testing.

Machine learning provides a variety of regression models, including k-nearest neighbors (KNN) [25], support vector regression (SVR) [26], random forests [27], and gradient boosting

trees [28,29]. For scenarios with high-dimensional features or strong nonlinearities, deep Neural Networks (ANNs) can learn hierarchical representations through multiple hidden layers and approximate complex nonlinear functions. In this work, a lightweight hybrid KNN-SVR model is adopted, which combines the local fitting capability of KNN with the global fitting capability of SVR. This model can achieve performance comparable to or even better than more complex models such as ANN and XGBoost, and is particularly well-suited for small sample and non-uniformly distributed datasets.

The main contributions of this work are as follows:

(1) We have found that the dynamic parameters of ADCs exhibit varying degrees of statistical correlation with INL*max_abs*

(2) A data-driven strategy for the indirect estimation of INL*max_abs* has been proposed, requiring only a small number of dynamic parameters as input features. Its effectiveness has been validated on datasets from five different types of ADCs.

(3) A hybrid KNN-SVR model is adopted, achieving performance comparable to or even better than complex models such as XGBoost and ANN. In addition, the proposed model exhibits more robust performance on datasets with different non-uniform distributions.

The remainder of this paper is organized as follows: Section 2 reviews existing strategies for optimizing INL testing. Section 3 primarily describes the implementation details of INL*max_abs* estimation using machine learning models; Section 4 presents and analyzes the experimental results; Section 5 summarizes the work of this paper.

## 2 Literature Review

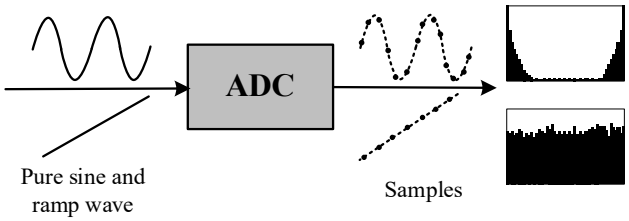



**Fig. 1** Standard INL testing method

As shown in Fig. 1, conventional INL testing typically uses a highly linear ramp signal or a low-distortion sinusoidal signal as the excitation, with statistical analysis performed using a code-density histogram [3]. In particular, the sine-wave histogram method yields a high concentration of samples near the waveform peaks and troughs. Consequently, more samples must be acquired to ensure sufficient counts for each output code and achieve high INL measurement accuracy.

Table 1 Comparison of Literature Survey

| **Authors & Methods** | **Approach** | **Advantages** | **Limitations** |
|---|---|---|---|
| Code-density histogram (IEEE [3]) | High-precision ramp/sinusoidal excitation → code-density histogram | High accuracy | Requires a large number of samples |
| Staircase-Wave Method (Satish [9]) | A single excitation waveform containing 32/64 DC steps | Compatible with the conventional definition of INL testing | Efficiency is limited by AWG memory capacity and data-transfer rate |
| Code-Selective Histogram (Zhao [10]) | Two-tone or multi-tone excitation signals | Increases the occurrence probability of critical codes | Covering all output codes still incurs a statistical cost |
| LCF/HCF Hybrid Testing (Serra [11]) | Multi-harmonic fitting for LCF; narrow-band triangular histogram fitting for HCF | Captures both the smooth trend and local variations of the INL curve | Requires two types of test excitation; HCF extraction still relies on histogram analysis |
| FSHT (Stefani [12]) | Reduced-sample SHT → initial INL→ low-pass/ENBW filtering for LCF extraction | Retains the standard SHT framework | Relies on low-frequency-dominant INL characteristics; high-frequency local errors may be lost |
| Wavelet-Based Reconstruction with Reduced Samples (Guo [13]) | Reduced-sample SHT → initial INL→ wavelet decomposition → thresholding and denoising → INL reconstruction | Suitable for high-accuracy ADCs | Still requires dedicated SHT and an initial INL estimate; sensitive to wavelet and threshold selection |

| | | | |
|---|---|---|---|
| Piecewise Polynomial/Maximum-Likelihood Method (Li [14], [15]) | INL segmentation → local polynomial modeling → maximum-likelihood parameter estimation | Structured and interpretable; suitable for high-resolution ADCs | Complex local variations may lead to model mismatch |
| Chebyshev-Based Method (Adamo [16]) | Sinusoidal excitation → harmonics → Chebyshev coefficients → INL reconstruction | Provides an explicit analytical relationship | Only suitable for smooth INL curves |
| Fourier-Spectrum-Based INL Estimation/BIST (Janik [17]; Kerzérho [18]) | Sinusoidal excitation →harmonics → Fourier-series coefficients → INL reconstruction; memory optimization for BIST | High INL curve fitting accuracy; suitable for on-chip integration | Difficult to accurately capture high-order harmonics |
| Data-driven indirect INL*max_abs* estimation (this work) | SNR/SINAD/THD/SFDR/ /harmonics → data-driven mapping → INL*max_abs* | Indirect estimation without direct INL testing. | Requires substantial training data |

To address the above issues, existing studies have improved the efficiency of INL testing from several perspectives, including excitation optimization, limited-sample data processing, and spectrum-based reconstruction. Table 1 summarizes the Approach, advantages, and limitations of representative methods.

In terms of excitation optimization, Satish et al. proposed a staircase-wave testing method, in which 32 or 64 DC levels are combined into a single test stimulus so that one acquisition contains information from multiple static test points [9]. Compared with applying the DC levels sequentially, this method shortens the test time by reducing the number of stimulus switching operations while preserving the physical meaning of static linearity testing. However, its efficiency is still constrained by the memory capacity of the arbitrary waveform generator and the waveform upload speed.

Zhao et al. proposed a code-selective histogram testing strategy [10]. This method uses two-tone or multi-tone excitation signals to increase the occurrence probability of critical codes and is suitable for SAR ADCs with well-defined critical-code structures. However, if a complete INL curve is required, the coverage and statistical accuracy of the remaining code regions must still be considered.

While retaining the histogram-based testing framework, Serra et al. divided INL into low code-frequency (LCF) and high code-frequency (HCF) components. The LCF component is obtained through multi-harmonic sinusoidal fitting, whereas the HCF component is extracted using a narrow-band triangular-wave histogram [11]. This method applies different test techniques to the smooth trend and local variations of the INL curve. However, its implementation requires two

different test excitations, and HCF estimation still relies on histogram statistics.

Stefani et al. further proposed the fast sine histogram test (FSHT), in which an initial INL sequence is first obtained from a reduced number of samples and is then appropriately filtered to suppress fluctuations caused by limited sampling [12]. This method retains the standard sine histogram testing framework, involves a relatively simple computational procedure, and can be readily integrated into conventional test flows. However, its effectiveness relies on the assumption that the INL curve is dominated by LCF components. When pronounced local discontinuities caused by HCF components are present, filtering may degrade the estimation of local details.

In 2023, Guo et al. proposed a high-accuracy ADC INL testing method based on wavelet decomposition and reconstruction [13]. The method first obtains an initial INL sequence from a reduced number of sine-histogram samples and then applies wavelet decomposition and threshold denoising to suppress random errors caused by limited sampling. However, an initial INL sequence must still be acquired, meaning that dedicated test excitation is not eliminated.

In addition to processing the initial INL sequence, structured modeling can also be used to reduce the required number of test samples. Li et al. successively investigated piecewise modeling with partial polynomial fitting and a piecewise polynomial method combined with maximum-likelihood estimation [14,15]. When the piecewise model can adequately represent the actual INL characteristics, relatively complete linearity information can be obtained using a reduced amount of test data. However, the estimation accuracy depends on the segmentation strategy and polynomial order, and model mismatch may occur for complex local discontinuities that are difficult to represent.

Another class of methods exploits the relationship between ADC static nonlinearity and harmonic distortion to reconstruct the INL curve indirectly from the output spectrum. Adamo et al. established a relationship between harmonic components and the coefficients of Chebyshev polynomials and used this relationship to estimate the INL curve [16]. This method provides a clear analytical relationship based on physical principles and can reduce the dependence on large-scale histogram statistics. However, a finite-order Chebyshev expansion is mainly suitable for relatively smooth INL curves and has limited capability to represent pronounced local discontinuities. In addition, FFT-based harmonic extraction requires careful control of spectral leakage caused by noncoherent sampling.

To improve the fitting accuracy for nonsmooth INL curves, Janik and Fresnaud employed a Fourier-series model to establish a mapping relationship between higher-order harmonics and the INL curve [17]. Kerzérho et al. further investigated model-parameter storage optimization for Fourier-series-based approaches in the context of built-in self-test (BIST) [18]. However, as higher reconstruction accuracy is required, reliably extracting a larger number of higher-order harmonics becomes increasingly challenging.

Most of the above methods still aim to obtain a complete or approximately complete INL curve. As the number of test samples is further reduced, random errors, filtering-induced smoothing, and model truncation may affect the recovery of local details. Therefore, a trade-off remains between the reconstruction accuracy of the complete INL curve and test efficiency.

In practical testing, INL is typically evaluated according to its maximum absolute value for pass/fail decisions. Therefore, an estimation method can be constructed directly around this target, without requiring reconstruction of the complete INL curve as an intermediate step. Based on this idea, this work establishes a data-driven mapping between the dynamic parameters and INL*max_abs* using SNR, SINAD, THD, SFDR, and selected harmonic components obtained from conventional dynamic testing. This strategy has a relatively low input dimensionality and has the potential to reduce the amount of dedicated data acquisition required for INL testing. Its effectiveness depends on the availability of sufficient labeled training data and appropriate test conditions.

# 3 The Implementation of INL*max_abs* Estimation

This section first presents the overall workflow for INL*max_abs* estimation, followed by a detailed elaboration on data acquisition, data processing, feature correlation analysis, and model construction.

## 3.1 Overall Flow

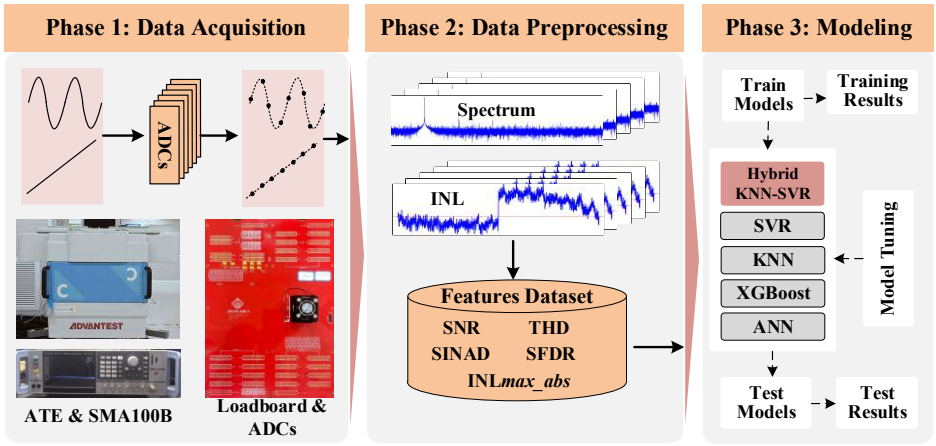



Fig 2 Overall Strategy of INL Estimation

Taking advantage of a large batch of ADC test samples, this paper establishes an indirect estimation model for the INL*max_abs*. The overall implementation procedure is presented in Fig. 2. First, an ADC test platform is established using Automated Test Equipment (ATE) to acquire the raw test data. Based on this platform, different excitation signals are applied to the ADC in accordance with the test specifications. Subsequently, the raw data from the static and dynamic responses are processed. The data processing procedure mainly includes raw data preprocessing, features extraction, and dataset partitioning. Finally, the features dataset is used to train different machine learning algorithms for the development of the INL*max_abs* estimation model. In addition, hyperparameter optimization is performed during model development to achieve the best estimation performance.

## 3.2 Data Acquisition

### 3.2.1 Test Equipment

The equipment setup for ADC testing is shown in Fig. 3. The system mainly consists of an ATE system, signal generator, device under test (DUT), and personal computer (PC). The ATE platform adopts the V93000 test system manufactured by Advantest. It integrates a power board, a digital board, a mixed-signal board, and a clock board to provide the power supply, sampling clock, and test excitation signals required for the ADC under test.

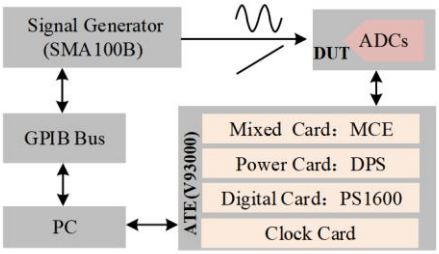



Fig. 3. Solution for ADCs testing

***Note:*** When the mixed-signal board in the ATE system cannot generate the high-purity sinusoidal signals required for ADC testing, a high-performance signal generator (e.g., the R&S® SMA100B RF and microwave generator) is controlled via the GPIB bus to supply the necessary test stimuli.

### 3.3.2 Test Conditions

To comprehensively validate the effectiveness of the proposed method, five different ADC devices were selected as experimental subjects. The device specifications, including manufacturer, resolution, architecture, and dataset size, are summarized in Table 2.

Table 2 Specifications of the ADCs under test

| Data set | Manufacturer | Resolution | Design Type | Dataset Size |
|---|---|---|---|---|
| AD0832 | TI | 8 | SAR | 682 |
| AD1674 | ADI | 12 | SAR | 1507 |
| AD7658 | ADI | 12 | SAR | 695 |
| AD9245 | ADI | 14 | Pipelined | 1187 |
| ADS8509 | TI | 16 | SAR | 697 |

The selected devices span resolutions from 8-bit to 16-bit and include both successive approximation register (SAR) and pipeline architectures. The devices are manufactured by Texas Instruments (TI) and Analog Devices Inc. (ADI). The number of collected dataset size ranges from 682 to 1507, which is sufficient for subsequent machine learning-based modeling.

Table 3 Dynamic Parameters Test Conditions

| Data set | Sampling rate | Excitation Type | Amplitude | Frequency | Samples |
|---|---|---|---|---|---|
| AD0832 | 250Ksps | Sine | Full Scale(0~5V) | 1.25KHz | 400 |
| AD1674 | 100Ksps | Sine | Full Scale (-10V~10V) | 10KHz | 2048 |
| AD7658 | 250Ksps | Sine | Full Scale (-10V~10V) | 50 KHz | 8192 |
| AD9245 | 50Msps | Sine | Full Scale ($2V_{PP}$) | 0.97MHz | 16384 |
| ADS8509 | 250Ksps | Sine | Full Scale(-10V~10V) | 20KHz | 8192 |

Table 4 Static parameters test conditions

| Data set | Sampling rate | Excitation Type | Amplitude | Frequency | Samples |
|---|---|---|---|---|---|
| AD0832 | 250Ksps | Ramp | 0~5V | 9.7KHz | 256×100 |
| AD1674 | 100Ksps | Sine | -10.05V~10.05V | 0.78KHz | 4096×256 |
| AD7658 | 250Ksps | Sine | -10.02V~10.02V | 1.95KHz | 4096×256 |
| AD9245 | 50Msps | Sine | $2.001V_{PP}$ | 0.97MHz | 16384×512 |
| ADS8509 | 250Ksps | Sine | -10.02V~10.02V | 0.97KHz | 65536×256 |

Tables 3 and 4 present the test conditions for the dynamic and static parameters of each ADC, respectively. It should be noted that dynamic parameter testing and static parameter testing are two independent test procedures, with different test objectives, excitation conditions, and data acquisition requirements. The main differences and characteristics of these two types of tests can be summarized as follows:

(1) **The amplitudes of the excitation signals are different.** In dynamic parameter testing, a near-full-scale sinusoidal signal is applied as the input excitation to adequately characterize the dynamic nonlinear behavior of the ADC over its input range. For static parameter testing based on the sine-wave histogram method, the amplitude of the sinusoidal signal is typically set slightly beyond the full-scale input range of the ADC to ensure coverage of the entire output code range. For example, the input ranges used for the static tests of the AD1674, AD7658, and ADS8509 are −10.05 to 10.05 V, −10.02 to 10.02 V, and −10.02 to 10.02 V, respectively.

(2) **The excitation frequencies used in both types of tests are relatively low with respect to the corresponding ADC sampling rates.** Using relatively low input frequencies helps mitigate the influence of high-frequency dynamic nonidealities, thereby enabling a more stable statistical relationship between the extracted dynamic parameters and INL*max_abs*. This provides more appropriate test conditions for the subsequent indirect estimation of INL*max_abs* from the dynamic parameters.

(3) **INL testing requires significantly more samples than dynamic parameter testing.** Dynamic parameters can generally be extracted through spectral analysis of an output sequence with a limited number of samples. In contrast, INL testing requires a sufficiently large number of code hits for the ADC output codes and therefore involves substantially more data acquisition. As shown in Tables 2 and 3, the static tests of the AD1674 and AD7658 each require approximately $4096\times256\approx1.05\times10^6$ samples, while the AD9245 requires approximately $16384\times512\approx8.39\times106$ samples. For the 16-bit ADS8509, the required number of samples further increases to $65536\times256\approx1.68\times10^7$. In comparison, the corresponding dynamic parameter tests generally require only hundreds to tens of thousands of samples. For example, the dynamic test of the ADS8509 requires only 8192 samples, whereas its static test requires approximately $1.68\times10^7$ samples, representing a difference of approximately 2048 times. Therefore, as the ADC resolution increases, the amount of data required for conventional INL testing increases substantially, resulting in greater test-time and data-processing overhead.

## 3.3 Data Processing

### 3.3.1 Preprocessing

As shown in Fig. 2, the preprocessing stage mainly processes the samples acquired by the ADC. First, a Fast Fourier Transform (FFT) is applied to the dynamic response data to obtain its spectrum. Second, the INL curve of the static response data is calculated using the code density histogram method.

### 3.3.2 Features Extraction

This subsection mainly focuses on extracting nonlinear parameters from the spectrum, which are subsequently used as input features for constructing the INL*max_abs* estimation model. When performing dynamic parameter testing, a pure analog sinusoidal signal must be applied to the ADC, as described by the following equation.

$$V_{in} = A_1 \sin(2\pi f_i t + \varphi) \quad (1)$$

Where $V_{in}$ represents continuous sinusoidal signal. After discrete sampling by the ADC, the output signal is represented as:

$$y_{out}[n] = A_0 + A_1 \sin\left(2\pi \frac{f_i}{f_s} n + \varphi_1\right) + \sum_{h=2}^{H} A_h \sin\left(2\pi \frac{h f_i}{f_s} n + \varphi_h\right) + \omega[n] \quad (2)$$

Here, $f_i$ and $f_s$ represent the input signal frequency and sampling frequency, respectively. $A_1$ denotes the fundamental amplitude, and $\varphi_1$ represents the initial phase. $A_h$ and $\varphi_h$ denote the amplitude and phase of the hth harmonic, respectively. $A_0$ represents the DC component. $\omega[n]$ represents noise, where $n = 0,1,\ldots,M-1$, and $M$ denotes the total number of samples.

Then, the FFT is applied to the M samples. In the spectrum plot, the relative power values of DC, fundamental, and harmonics are represented as:

$$HD_h = 10log_{10}\left(\frac{{A_h}^2}{{A_1}^2}\right) \qquad h = 0,1..n \quad (3)$$

Where, $HD_0$ denotes the power noise of the DC offset, $HD_1$ is the relative value of the fundamental power, typically taken as 1; when $h = 2,3,\ldots,n$, $HD_h$ represents the power of the $h$-th harmonic power.

The remaining spectrum parameters such as SNR, SINAD, THD, and SFDR are calculated according to the following formula:

$$SNR = 10log_{10}\left(\frac{A_1^2/2}{P_{noise}}\right) \quad (4)$$

$$SINAD = 10log_{10}\left(\frac{A_1^2/2}{\sum_{h=2}^{H} A_h^2/2 + P_{noise}}\right) \quad (5)$$

$$THD = 10log_{10}\left(\frac{\sum_{h=2}^{H} A_h^2}{A_1^2}\right) \quad (6)$$

$$SFDR = 10log_{10}\frac{A_1^2}{max(A_h^2)} \quad (7)$$

$P_{noise}$ represents the sum of the power at all remaining frequency points, excluding the fundamental wave, harmonics, and DC components.

In summary, the dynamic response of an ADC contains abundant nonlinear characteristic parameters, including SNR, SINAD, THD, SFDR, noise power, DC noise, and harmonic components. Among these parameters, SNR reflects the noise power relative to the fundamental signal power and is used to characterize the noise level of the ADC. THD represents the overall harmonic power relative to the fundamental signal power and characterizes the distortion level of the ADC. SINAD jointly accounts for the effects of noise and harmonic distortion, while SFDR measures the ADC's ability to suppress spurious and harmonic components. In contrast, INL*max_abs* quantifies the maximum deviation of the ADC's actual static transfer characteristic from the ideal transfer characteristic. Although the dynamic parameters and INL*max_abs* characterize ADC nonlinearity from the perspectives of dynamic and static responses, respectively, both are affected by the nonideal nonlinear behavior of the ADC. Therefore, statistical correlations may exist between them, which will be analyzed in detail in Section 3.4.

#### 3.3.3 Dataset Partitioning

To ensure a fair evaluation of the model performance, 30% of the samples were randomly selected from the dataset as an independent test set. These test samples were completely excluded from all subsequent model training and validation processes. The remaining 70% of the data were subjected to five-fold cross-validation, where one subset was sequentially used as the validation set and the other four subsets were used for model training. This process was repeated five times to complete the training and validation procedure. The optimal hyperparameters of the model are selected based on the results of five validation, and its generalization performance is subsequently evaluated on an independent test set.

### 3.4 Features Analysis

To further investigate the correlation between INL*max_abs* and dynamic parameters, this section employs the Pearson correlation coefficient and two-dimensional scatter plots to conduct analysis across different ADC datasets and to validate the feasibility of the proposed method.

#### 3.4.1 Correlation Analysis of Features

The Pearson correlation coefficient is used to measure the strength of the linear relationship between two variables. Its formula is given as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{8}$$

Where $X_i$ and $Y_i$ denote the *i*-th sample values of the two variables, respectively; $\bar{X}$ and $\bar{Y}$ denote the corresponding sample means; and $n$ is the total number of samples. The correlation coefficient $r$ ranges from -1 to 1, with values closer to 1 indicating a stronger correlation between the two variables. Specifically, $r>0$ indicates a positive correlation, whereas $r < 0$ indicates a negative correlation.

The strength of the correlation was interpreted according to the following criteria:

- $|r|> 0.7$: Strong correlation
- $0.4 \leq |r| \leq 0.7$: Moderate correlation
- $|r| < 0.4$: Weak correlation

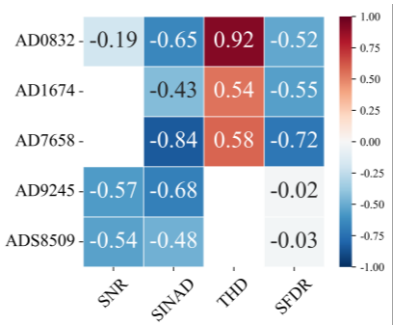



Fig. 4. Pearson correlation coefficients between INL*max_abs* and dynamic parameters in different ADC datasets

**Note:** Some feature data are unavailable due to they were not preserved during testing, and the corresponding measurements could not be repeated.

Fig. 4 presents the Pearson correlation coefficients between dynamic parameters and INL*max_abs* for different datasets. From the sign of the correlation coefficients, THD exhibit positive correlation with INL*max_abs*, whereas SNR, SINAD, and SFDR exhibit negative correlation. This indicates that as INL*max_abs* increases, SNR, SINAD, and SFDR decrease,

whereas THD increases. These observations are consistent with the nonlinear behavior of ADCs. Specifically, deterioration of INL*max_abs* indicates increased harmonic distortion and noise within the ADC, resulting in degraded dynamic nonlinear performance.

Next, the strength of the correlations between the dynamic parameters and INL*max_abs* was analyzed. For the AD0832 dataset, THD exhibits a strong correlation with INL*max_abs* ($r$ = 0.92), whereas SINAD and SFDR exhibit moderate correlations ($r$ = -0.65 and -0.52, respectively). In contrast, SNR shows a weak correlation ($r$ = -0.19). For the AD1674 dataset, SINAD, THD, and SFDR all exhibit moderate correlations with INL*max_abs*, with correlation coefficients of -0.43, 0.58, and -0.55, respectively. For the AD7658 dataset, SINAD and SFDR exhibit strong correlations ($r$ = -0.84 and -0.72, respectively), whereas THD exhibits a moderate correlation ($r$ = 0.58). For the AD9245 dataset, SNR and SINAD exhibit moderate correlations ($r$ = -0.57 and -0.68, respectively), whereas SFDR exhibits a weak correlation ($r$ = -0.02). For the ADS8509 dataset, SNR and SINAD exhibit moderate correlations ($r$ = -0.54 and -0.48, respectively), whereas SFDR exhibits a weak correlation ($r$ = -0.03).

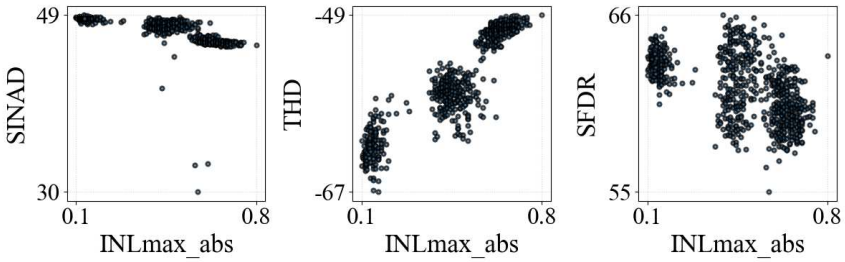
Overall, since SINAD jointly accounts for the effects of noise and harmonic distortion on the dynamic performance of an ADC, it exhibits a moderate or strong linear correlation with INL*max_abs* across all datasets. THD characterizes the overall level of harmonic distortion and also shows a moderate or strong correlation in all four datasets. In contrast, SFDR primarily characterizes the largest spurious component relative to the fundamental component, and its correlation with INL*max_abs* varies considerably among the datasets. Specifically, SFDR exhibits a moderate or strong correlation in the AD0832, AD1674, and AD7658 datasets, whereas only a weak correlation is observed in the AD9245 and ADS8509 datasets. Similarly, SNR exhibits different correlation strengths across the three datasets.

These results demonstrate that INL*max_abs* exhibits different degrees of statistical correlation with different dynamic parameters, and that these correlations are not consistent across all ADC datasets. Therefore, a single dynamic parameter may not be sufficient to consistently and comprehensively characterize the variation in INL*max_abs*. By combining multiple dynamic parameters, including SNR, SINAD, THD, SFDR, and individual harmonic components, statistical information associated with INL*max_abs* can be captured from multiple perspectives. These observations provide a statistical basis for subsequently using multiple dynamic parameters as

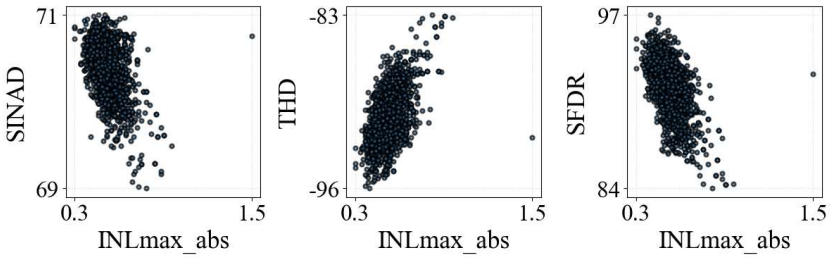
input features to establish a data-driven model for the indirect estimation of INL*max_abs*.

### 3.4.2 Distribution of Dynamic Parameters

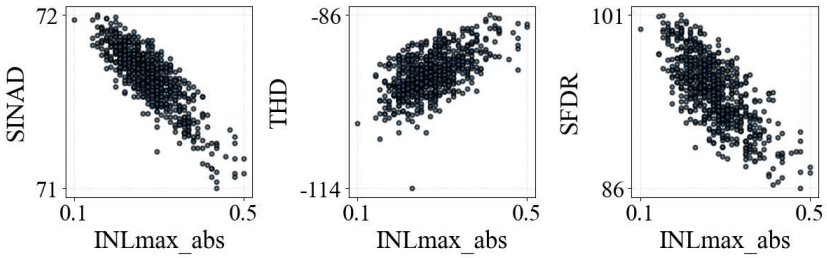
Scatter plots provide an intuitive visualization of the relationships between the dynamic parameters and INL*max_abs*, complementing the Pearson correlation analysis. As shown in Fig. 5, the three features with the largest absolute Pearson correlation coefficients were selected for each dataset, and their scatter plots were generated.



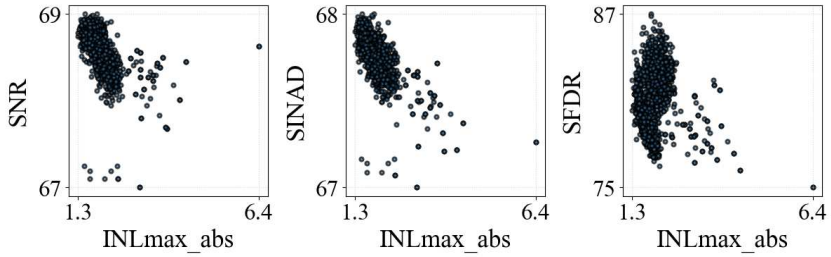

(a) AD0832



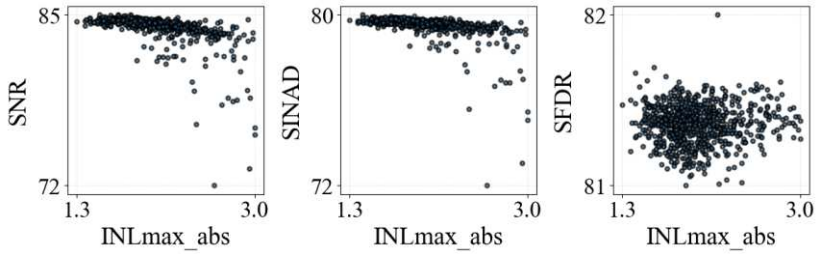

(b) AD1674




(c) AD7658




(d) AD9245




(e) ADS8509

Fig. 5 Distribution of dynamic parameters vs INL*max_abs*

From the scatter plot trends, SNR, SINAD, and SFDR generally decrease as INL*max_abs* increases, whereas THD exhibit increasing trends. These trends are consistent with the signs of the Pearson. correlation coefficients. The scatter plots also illustrate the relationship between the magnitude of the Pearson correlation coefficient and the degree of linear dependence. In general, a larger $|r|$ corresponds to a stronger linear correlation, resulting in a more concentrated distribution of data points and a more pronounced linear trend. For example, THD in the AD0832 dataset, as well as SINAD and SFDR in the AD7658 dataset, exhibit strong correlations with INL*max_abs*. Their scatter plots display clear linear trends with tightly clustered data points, indicating that these parameters are highly sensitive to variations in INL*max_abs*. By comparison, SINAD, THD, and SFDR in the AD1674 dataset, exhibit moderate correlations with INL*max_abs*. Although clear increasing or decreasing trends remain observable, the scatter points are relatively more dispersed. When the correlation is weak, the linear relationship becomes less apparent, and the scatter points are distributed more broadly, as exemplified by SFDR in the ADS8509 dataset.

### 3.5 Hybrid KNN-SVR model

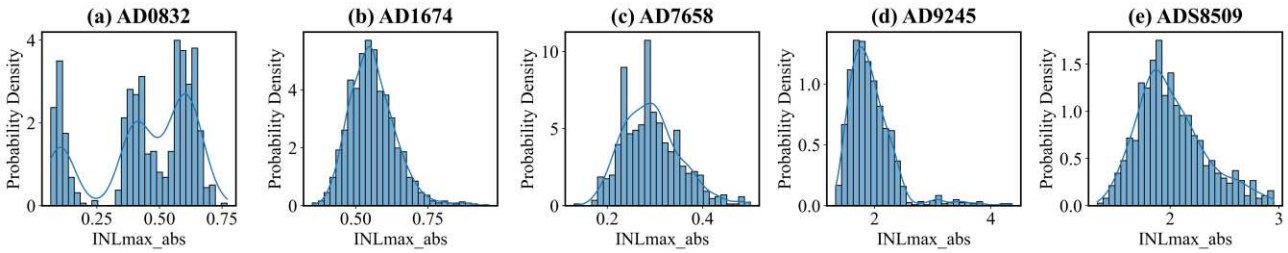



Fig 6 Distribution of INL*max_abs* across different datasets

As shown in Fig. 6, the distribution characteristics of INL*max_abs* vary considerably across the different datasets. Among them, only the AD0832 dataset exhibits an approximately uniform distribution, whereas the other four datasets show different degrees of non-uniformity. Specifically, the distributions of the AD1674 and AD7658 datasets approximately resemble normal distributions, while those of the AD9245 and ADS8509 datasets exhibit a certain degree of skewness. In addition, the number of samples in these datasets ranges from 682 to 1507, indicating that they can generally be regarded as small- to medium-sized datasets. These results demonstrate considerable differences among the ADC datasets in terms of both sample distribution characteristics and dataset size, making it difficult for a single regression model to consistently achieve optimal estimation performance across all datasets.

To overcome these limitations, a robust Hybrid KNN-SVR model is proposed. The proposed model dynamically fuses the results of two complementary base models, KNN and SVR, according to the local sample density of input features in the feature space. The implementation procedure of the proposed Hybrid KNN-SVR algorithm is summarized in Table 5, and its key steps are described in detail below.

KNN is a typical lazy learning algorithm that does not require an explicit training process. For a test sample, the estimation is obtained by identifying its *K* nearest neighbors in the training set and averaging their corresponding target values, which can be expressed as：

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i \tag{9}$$

KNN possesses strong local fitting capability. In densely populated regions, reliable neighboring samples can be identified, leading to accurate estimation. However, in sparse regions,

insufficient neighbors may result in unreliable estimates. Consequently, the performance of KNN is highly dependent on the data distribution, making it difficult to maintain consistent estimation accuracy across non-uniform datasets.

Unlike KNN, SVR learns the global mapping between the input features and the target variable by constructing an optimal regression hyperplane. As a result, SVR can provide stable estimation even for samples located in sparse regions. However, because SVR primarily captures the global trend of the data, its ability to model local variations in high-density regions may be limited. In contrast, KNN can better exploit the local neighborhood information in such regions, resulting in stronger local fitting capability.

Table 5 Hybrid KNN-SVR algorithm

**Algorithm:** Hybrid KNN-SVR Model

**Input:** Sample data (*x, y*), split into training and test sets with a 7:3 ratio.

*k* (number of neighbors) and $\beta$ (steepness factor of the sigmoid function).

1. Train KNN and SVR models on (*x, y*) to obtain $f_{\text{KNN}}$ and $f_{\text{SVR}}$, respectively.

2. Compute the local density for each sample:

$$\delta(x_i) = 1/avg_dist(x_i, k\ nearest\ neighbors)$$

3. Compute the mixing coefficient:

$$\alpha_i = 1/(1 + exp(-\beta \cdot \delta(x_i)))$$

4. Compute the final hybrid prediction estimation:

$$\hat{y}_i^{\text{KNN}} = f_{\text{KNN}}(x_i), \hat{y}_i^{\text{SVR}} = f_{\text{SVR}}(x_i)$$

5. Final estimation：

$$\hat{y}_i^{\text{hybrid}} = \alpha_i \cdot \hat{y}_i^{\text{KNN}} + (1 - \alpha_i) \cdot \hat{y}_i^{\text{SVR}}$$

**Output:** Final estimated value $\hat{y}_i^{\text{hybrid}}$

To exploit the complementary strengths of these two models, the proposed Hybrid KNN-SVR model introduces an adaptive weighting coefficient, $\alpha_i$, to dynamically combine the local fitting capability of KNN with the global generalization capability of SVR. Specifically, KNN is assigned a larger weight in densely populated regions, whereas SVR is given greater emphasis in sparse regions.

The implementation procedure is summarized in Table 4.

For an input feature vector $x_i$, the estimation of KNN and SVR are first obtained as:

$$\hat{y}_i^{\text{KNN}} = f_{\text{KNN}}(x_i), \hat{y}_i^{\text{SVR}} = f_{\text{SVR}}(x_i) \quad (10)$$

Then, the final hybrid estimation is calculated as:

$$\hat{y}_i^{\text{hybrid}} = \alpha_i \cdot \hat{y}_i^{\text{KNN}} + (1 - \alpha_i) \cdot \hat{y}_i^{\text{SVR}} \tag{11}$$

The adaptive weighting coefficient $\alpha_i$ is determined by a sigmoid function,

$$\alpha_i = \frac{1}{1 + exp\left(-\beta \cdot \delta(x_i)\right)} \tag{12}$$

Where $\delta(x_i)$ measures the local sample density around $x_i$, which is typically defined as the inverse of the average Euclidean distance from $x_i$ to its k nearest neighbors. A larger value of $\delta(x_i)$ indicates that the sample is located in a densely populated region, whereas a smaller value corresponds to a sparse or boundary region.

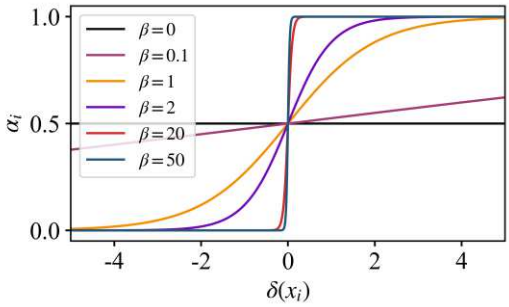



Fig. 7 Sigmoid functions with different values of $\beta$

As illustrated in Fig. 7, the relationship between $\alpha_i$ and $\delta(x_i)$ is established through a sigmoid function. The hyperparameter $\beta$ controls the steepness of the sigmoid transition.

- When $\beta \to \infty$, the sigmoid function degenerates into a unit step function, which yields hard-switching behavior. Samples with $\delta(x_i)$ above the threshold are estimated exclusively by KNN, whereas those below the threshold rely entirely on SVR, with no intermediate transition.
- When $\beta = 0$, $\alpha_i$=0.5 meaning both models contribute equally regardless of sample density.
- For moderate values of $\beta$, the sigmoid function provides a smooth transition, and the

contributions of KNN and SVR vary gradually with sample density.

In addition, this paper also compares the proposed model with KNN, SVR, XGBoost, and ANN models. The architectures and hyperparameter ranges of these models are listed in Table 6.

The hardware and software configuration used for training the experimental model is as follows: an RTX 3060 Laptop GPU; Python 3.11.15; XGBoost 2.1.1; NumPy 1.26.4; Keras 2.13.1; and SHAP 0.47.2.

Table 6 Structures and parameters of different models

| Algorithms | Parameters | Value range |
|---|---|---|
| SVR | Kernel | Rbf |
| | C | [0.1, 0.5, 1, 2, 5, 10, 20, 50, 100] |
| | epsilon | [0.001, 0.01, 0.05, 0.1, 0.2, 0.5] |
| | Gamma | ['scale', 'auto', 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.5, 1, 2] |
| KNN | Number of neighbors | [3,5,7,9,11,13,15,17,19,21] |
| | Weights | ['uniform', 'distance'] |
| XGBoost | Number of decision trees | [20, 50, 100, 150, 200] |
| | Maximum depth | [2, 3, 4, 5] |
| | Learning rate | [0.01, 0.05, 0.1, 0.2] |
| | Colsample bytree | [0.6, 0.8, 1.0] |
| | Subsample | [0.6, 0.8, 1.0] |
| | Reg alpha | [0.001, 0.01, 0.1] |
| | Reg lambda | [0.001, 0.01, 0.1] |
| ANN | Number of nodes in input layer | Number of input features |
| | Number of nodes in 1st hidden layer | Layer size |
| | Number of nodes in 2nd hidden layer | Layer size |
| | Number of nodes in 3rd hidden layer | Layer size |
| | Layer size | [32, 64, 128, 256] |
| | Activation function | Relu |
| | Learning rate | [1e-4, 5e-4, 1e-3, 5e-3] |
| | Dropout | [0.1, 0.2, 0.3, 0.5] |

### 3.6 Evaluation Metrics

To quantify the error between the true values and the estimated values, and to evaluate the

model's fitting performance, this work employs the coefficient of determination $R^2$, root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) to establish a comprehensive evaluation system. Their calculation formulas are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{13}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{16}$$

Where $\hat{y}_i$ represents the estimated value, $y_i$ indicates the true value, $\bar{y}_i$ denotes the mean of the true INL*max_abs* values, and $n$ is the number of samples in the dataset.

$R^2$ is the most commonly used goodness-of-fit indicator, representing the proportion of the variance in the dependent variable that can be explained by the model. $R^2$ ranges from 0 to 1, and the closer the value is to 1, the better the goodness-of-fit of the model. RMSE and MAE quantify the error accuracy between measured and estimated values, with the same dimension as the estimated variable; MAPE quantifies the estimation error as a percentage relative to the measured values. The closer these three metrics are to 0, the smaller the model's estimation error.

# 4 Results and Discussion

This section primarily presents the results and discusses the factors that influence the estimation outcomes, such as input features and models.

## 4.1 Results

### 4.1.1 Metric Results

First, Table 7 presents the performance of the five machine learning algorithms on different test datasets. For the subsequent analysis, the results with the best RMSE performance in each dataset are selected for discussion.

From the perspective of goodness-of-fit, the $R^2$ values for all five test datasets are greater than 0.4. Specifically, the $R^2$ values for datasets AD0832, AD7658, and AD9245 are 0.9294, 0.7538, and 0.7336, respectively, all exceeding 0.7 and indicating excellent goodness-of-fit. The remaining datasets AD1674 and ADS8509, yield $R^2$ values of 0.4229 and 0.5080, respectively, demonstrating acceptable performance.

In terms of the MAPE, the performance across the test datasets is approximately 10%. The values for datasets AD1674, AD7658, AD9245, and ADS8509 are 8.86%, 8.22%, 7.24%, and 7.80%, respectively, all below 10%. In comparison, datasets AD0832 exhibit relatively poorer values of 15.38%.

RMSE and MAE are used to quantify the errors between the estimated and measured values of INL*max_abs*, with both metrics expressed in units of LSB. Their values are influenced by various factors, including the resolution of ADC and the distribution range of INL*max_abs*. Therefore, these metrics are not directly comparable across different datasets. The RMSE values for the five datasets are 0.0532, 0.0607, 0.0301, 0.2006, and 0.2083, respectively, while the corresponding MAE values are 0.0385, 0.0486, 0.0235, 0.1423, and 0.1546.

Table 7：Estimation results of INL*max_abs* in different models

| Dataset | Models | R2 | RMSE (LSB) | MAE (LSB) | MAPE (%) |
|---|---|---|---|---|---|
| AD0832 | SVR | 0.9175 | 0.0575 | 0.0423 | 18.02 |
| | **KNN** | 0.9229 | 0.0556 | 0.0386 | 14.77 |
| | XGBoost | 0.8953 | 0.0648 | 0.0425 | 18.38 |
| | ANN | 0.9096 | 0.0602 | 0.0433 | 15.72 |
| | **Hybrid KNN-SVR** | **0.9294** | **0.0532** | **0.0385** | **15.38** |
| AD1674 | SVR | 0.4142 | 0.0612 | 0.0487 | 8.84 |
| | KNN | 0.3893 | 0.0625 | 0.0495 | 9.01 |
| | **XGBoost** | 0.4167 | 0.0611 | 0.0493 | 9.03 |
| | ANN | 0.3939 | 0.0623 | 0.0501 | 9.14 |
| | **Hybrid KNN-SVR** | **0.4229** | **0.0607** | **0.0486** | **8.86** |
| AD7658 | **SVR** | 0.7501 | 0.0304 | 0.0240 | 8.51 |
| | KNN | 0.7381 | 0.0311 | 0.0242 | 8.63 |
| | XGBoost | 0.729 | 0.0316 | 0.0244 | 8.67 |

| | | | | | |
|---|---|---|---|---|---|
| | ANN | 0.7316 | 0.0315 | 0.0243 | 8.57 |
| | **Hybrid KNN-SVR** | **0.7538** | **0.0301** | **0.0235** | **8.22** |
| AD9245 | SVR | 0.6424 | 0.2140 | 0.1432 | 7.30 |
| | KNN | 0.6918 | 0.2158 | 0.1578 | 8.02 |
| | **XGBoost** | 0.7251 | 0.2038 | 0.1509 | 7.74 |
| | ANN | 0.6211 | 0.2203 | 0.1531 | 7.78 |
| | **Hybrid KNN-SVR** | **0.7336** | **0.2006** | **0.1423** | **7.24** |
| ADS8509 | SVR | 0.4446 | 0.2213 | 0.1639 | 8.27 |
| | KNN | 0.4878 | 0.2125 | 0.1561 | 7.80 |
| | **XGBoost** | **0.5080** | **0.2083** | **0.1546** | **7.83** |
| | **ANN** | **0.4965** | **0.2107** | **0.1561** | **7.80** |
| | **Hybrid KNN-SVR** | **0.4909** | **0.2119** | **0.1553** | **7.79** |

### 4.1.2 Models Comparison

As shown in Table 7, taking the RMSE metric as an example, conventional machine learning models exhibit substantial performance variations across different datasets. Among the five datasets, the best-performing models are KNN, XGBoost, SVR, XGBoost, and XGBoost, respectively. These differences are mainly attributed to factors such as sample distribution, sample size, feature importance, and the presence of noisy data.

The proposed Hybrid KNN-SVR model combines the local fitting capability of KNN with the global fitting capability of SVR. The experimental results show that: (1) the proposed model consistently outperforms the two base models, KNN and SVR, across all datasets; and (2) on the AD0832, AD1674, AD7658, and AD9245 datasets, the Hybrid KNN-SVR model also achieves better performance than XGBoost and ANN. Only on the ADS8509 dataset does its performance remain comparable to that of XGBoost and ANN. Overall, the Hybrid KNN-SVR model demonstrates strong robustness across different ADC datasets.

### 4.1.3 Results Distribution

To further validate the reliability and effectiveness of the proposed model, scatter plots and probability distribution plots were employed to visually evaluate the estimation results for both the training and test sets. As shown in Fig. 8, the central panel of each subplot presents the scatter plot comparing the estimated and measured INL*max_abs* values, while the top and right panels

illustrate the probability distributions of the measured and estimated values, respectively.

The probability distribution of the measured values indicates that the training and test sets follow similar data distributions, thereby avoiding systematic bias caused by distribution discrepancies. Moreover, the probability distributions of the estimated values exhibit good agreement with those of the measured values for both the training and test sets, demonstrating that the proposed model possesses strong generalization capability without showing signs of overfitting.

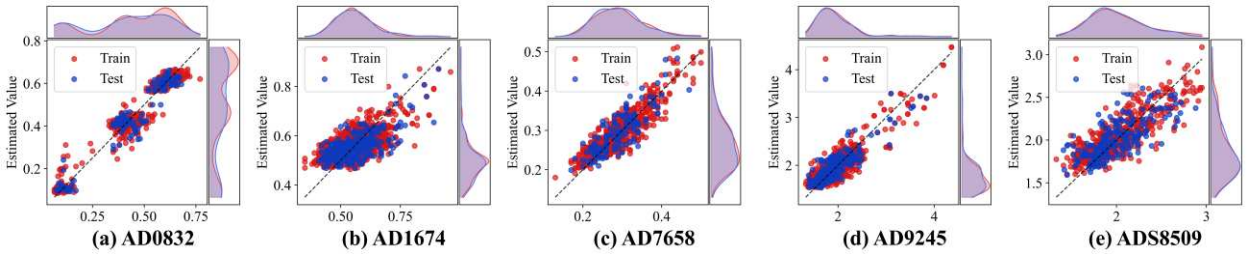



Fig. 8 Scatter plot of the estimation results for different datasets

The scatter plots further reveal that most data points are closely clustered around the ideal fitting line. In general, a higher $R^2$ value corresponds to a tighter concentration of scatter points around the ideal line, as observed for datasets AD0832, AD7658, and AD9245. In contrast, for datasets with relatively lower $R^2$ values, such as ADS6445, AD1674, and ADS8509, the scatter points are more dispersed around the ideal fitting line.

## 4.2 Features Importance Analysis

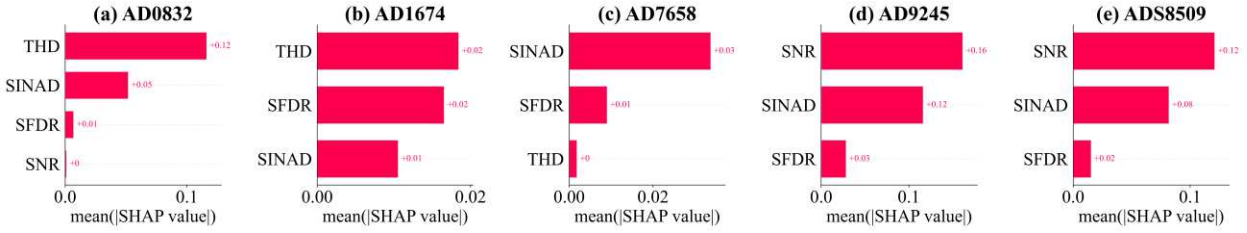



Fig. 9 Features importance analysis

To further investigate the contributions of different input features to the estimation results,

SHAP analysis was performed for each dataset, as illustrated in Fig. 9. The results reveal that the importance of dynamic parameters varies considerably across different datasets, indicating that the nonlinearity formation mechanisms differ substantially among ADC devices.

Specifically, for the AD0832 dataset, THD and SINAD are the dominant features, whereas SFDR exhibits relatively low importance and SNR contributes negligibly. For AD1674, the feature importance decreases sequentially from THD to SFDR and SINAD, with these three features exhibiting comparable contributions. In the AD7658 dataset, SINAD is considerably more important than both SFDR and THD. The AD9245 and ADS8509 datasets exhibit similar patterns, in which SNR and SINAD are the two most influential features, while SFDR contributes relatively little.

Overall, SINAD, which jointly characterizes the effects of noise and harmonic distortion, consistently exhibits high feature importance across the different ADC datasets. In contrast, the contributions of SNR, THD, and SFDR vary considerably among the datasets. As indicated by Eq. (2), the dynamic nonlinear performance of an ADC is jointly associated with noise and harmonic distortion. Furthermore, owing to differences in the nonlinear characteristics of different ADCs, the statistical relationships between these dynamic parameters and INL*max_abs* may vary across devices. Notably, the feature-importance patterns shown in Fig. 9 are generally consistent with the Pearson correlation results presented in Fig. 4, further supporting the use of multiple dynamic parameters as complementary input features for the indirect estimation of INL*max_abs*.

### 4.3 Discussion of Contributions and Limitations

(1) Contributions

This work proposes a data-driven method for indirectly estimating INLmax_abs, achieving RMSE values of 0.0532, 0.0607, 0.0301, 0.2006, and 0.2083 LSB across five datasets, respectively. Once trained and validated, the model estimates INL*max_abs* solely from the dynamic parameters of the device under test, without requiring an additional dedicated static-test stimulus. Given the sampling requirements reported in Tables 3 and 4, this strategy can reduce dedicated INL data acquisition and thereby improve test efficiency. Furthermore, the experimental results demonstrate that the Hybrid KNN–SVR model provides good robustness across small and medium-sized

datasets with different distributions.

(2) Limitations

First, the applicability of the proposed method is subject to constraints on the test stimulus. As shown in Tables 3 and 4, both static and dynamic tests in this work use low-frequency stimuli covering the full-scale input range. Low-frequency excitation helps reduce the influence of frequency-dependent dynamic nonlinearity on the estimation relationship, while full-scale excitation helps characterize nonlinearity over the entire input range. This is consistent with the approach of Duan et al. [24], who estimated low-frequency spectral performance from measured INL values. Their study established an analytical relationship between INL and harmonic distortion, enabling the estimation of THD and harmonic-dominated SFDR, but could not capture dynamic effects or nonharmonic spurs. Therefore, the effectiveness of the proposed method at higher input frequencies or under non-full-scale excitation requires further validation.

Second, reliable indirect estimation of DNL has not yet been achieved, and INL*max_abs* estimation cannot replace missing-code detection. At this stage, the method is therefore better suited as a supplementary estimation approach within production-test flows for well-validated manufacturing processes. Dedicated DNL or missing-code testing remains necessary when the product requires a no-missing-code guarantee.

Finally, the five ADCs used for validation have sampling rates of 250 Ksps, 100 Ksps, 250 Ksps, 50 Msps, and 250 Ksps, respectively, with a maximum resolution of 16 bits. The sampling-rate and resolution ranges covered by the present results are therefore limited, and applicability to faster or higher-resolution ADCs requires further validation.

# 5 Conclusion

This paper proposes a data-driven indirect estimation method for INL*max_abs*, using dynamic parameters as input features. The proposed method has been validated on datasets from five different ADCs, covering resolutions of 8-bit, 12-bit, 14-bit, and 16-bit, as well as both SAR and Pipelined architectures. The results show that the MAPE values are all around 10%. In terms of model performance, the Hybrid KNN-SVR model consistently exhibits stable performance across small-to medium-sized datasets with varying distributions. In future work, we will enhance

feature extraction from spectral sequences to improve INL*max_abs* estimation accuracy. We will also investigate how the model can rapidly adapt to samples from different batches during production testing

**Authors' Contributions** Rongjie Li, PhD, was responsible for the experimental data analysis and manuscript preparation. Prof. Houjun Wang provided guidance on the experiments. Pu Pu and Jia Li assisted with experimental data collection. Associate Professors Guibing Zhu and Wanyu Yang reviewed the manuscript.

**Funding** The authors declare that no funds were received during the preparation of this manuscript.

**Data Availability** The raw data used in this study were provided by the Integrated Circuit Testing Laboratory of the University of Electronic Science and Technology of China. However, access to these data is restricted. As the data were used with permission for this study, they are not publicly available.

## References

[1] Huang L, Song T (2022) VLSI test through an improved LDA classification algorithm for test cost reduction. Microelectron J 125:105461. https://doi.org/10.1016/j.mejo.2022.105461

[2] Song T, Huang Z, Guo X (2022) RLDA: Valid test pattern identification by machine learning classification method for VLSI test. Microelectron J 128:105549. https://doi.org/10.1016/j.mejo.2022.105549

[3] IEEE (2011) IEEE standard for terminology and test methods for analog-to-digital converters. IEEE Std 1241-2010 (Revision of IEEE Std 1241-2000), pp 1–139. https://doi.org/10.1109/IEEESTD.2011.5692956

[4] Alegria F, Arpaia P, Da Cruz Serra AM, Daponte P (2002) Performance analysis of an ADC histogram test using small triangular waves. IEEE Trans Instrum Meas 51:723–729. https://doi.org/10.1109/TIM.2002.803292

[5] Huang Y et al (2024) A high constancy and noise suppression voltage shift generator in SEIR-based BIST circuit for ADC linearity test. Microelectron J 151:106341. https://doi.org/10.1016/j.mejo.2024.106341

[6] Palfi V (2019) An improved sine wave histogram test method for ADC characterization. IEEE

Trans Instrum Meas 68:3446–3455. https://doi.org/10.1109/TIM.2018.2878593

[7] Vora SC, Satish L (2011) ADC static nonlinearity estimation using linearity property of sinewave. IEEE Trans Instrum Meas 60:1283–1290. https://doi.org/10.1109/TIM.2010.2084773

[8] Chen T et al (2020) A low-cost on-chip built-in self-test solution for ADC linearity test. IEEE Trans Instrum Meas 69:3516–3526. https://doi.org/10.1109/TIM.2019.2936716

[9] Satish L, Vora SC, Sinha AK (2005) A time efficient method for determination of static non-linearities of high-speed high-resolution ADCs. Measurement 38:77–88. https://doi.org/10.1016/j.measurement.2005.06.003

[10] Zhao Y, Katoh K, Kuwana A et al (2022) Revisit to histogram method for ADC linearity test: examination of input signal and ratio of input and sampling frequencies. J Electron Test 38:21–38. https://doi.org/10.1007/s10836-022-05988-y

[11] Serra AC, da Silva MF, Ramos PM, Martins RC, Michaeli L, Saliga J (2005) Combined spectral and histogram analysis for fast ADC testing. IEEE Trans Instrum Meas 54(4):1617–1623. https://doi.org/10.1109/TIM.2005.851057

[12] Stefani F, Macii D, Moschitta A, Carbone P, Petri D (2006) Simple and time-effective procedure for ADC INL estimation. IEEE Trans Instrum Meas 55(4):1383–1389. https://doi.org/10.1109/TIM.2006.876395

[13] Guo X, Yan X, Ma S et al (2023) Wavelet decomposition and reconstruction-based INL method for testing high-precision ADC. Microelectron J 142:105945. https://doi.org/10.1016/j.mejo.2023.105945

[14] Li D, Zhu Y, Wang L, Liu S, Zhu Z (2024) Low-cost linearity testing of high-resolution ADCs using segmentation modeling and partial polynomial fitting. In: Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS), pp 1–4. https://doi.org/10.1109/ISCAS58744.2024.10558217

[15] Li D, Zhu Y, Wang L, Liu S, Zhu Z (2025) Cost-efficient ADC linearity testing based on maximum likelihood partial polynomial segmentation method. IEEE Trans Instrum Meas 74:1–12. https://doi.org/10.1109/TIM.2025.3629852

[16] Adamo F, Attivissimo F, Giaquinto N, Savino M (2002) FFT test of A/D converters to determine the integral nonlinearity. IEEE Trans Instrum Meas 51(5):1050–1054.

https://doi.org/10.1109/TIM.2002.807795

[17] Janik JM, Fresnaud V (2007) A spectral approach to estimate the INL of A/D converter. Comput Stand Interfaces 29(1):31–37. https://doi.org/10.1016/j.csi.2005.12.004

[18] Kerzérho V, Bernard S, Cauvet P, Janik JM (2006) A first step for an INL spectral-based BIST: the memory optimization. J Electron Test 22(4–6):351–357. https://doi.org/10.1007/s10836-006-0186-z

[19] Rusen V, Krukonis A, Plonis D (2021) Prediction of parameters of semiconductor band-pass filters using artificial neural network. In: Proceedings of the 2020 IEEE 8th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), Vilnius, Lithuania, pp 1–4. https://doi.org/10.1109/AIEEE51419.2021.9435748

[20] Roy S, Millican SK, Agrawal VD (2021) Special session—Machine learning in test: A survey of analog, digital, memory, and RF integrated circuits. In: Proceedings of the 2021 IEEE 39th VLSI Test Symposium (VTS), San Diego, CA, USA, pp 1–14. https://doi.org/10.1109/VTS50974.2021.9441051

[21] Lee J, Walker DMH, Milor L, Peng Y, Hill G (1999) IC performance prediction for test cost reduction. In: Proceedings of the 1999 IEEE International Symposium on Semiconductor Manufacturing Conference, Santa Clara, CA, USA, pp 111–114. https://doi.org/10.1109/ISSM.1999.808750

[22] Hou L, Liu Y, Xie W, Dai Z, Yang W, Zhao Y (2022) Statistical neural network (SNN) for predicting signal-to-noise ratio (SNR) from static parameters and its validation in 16-bit, 125-MSPS analog-to-digital converters (ADCs). Rev Sci Instrum 93:084701. https://doi.org/10.1063/5.0093709

[23] Duan J, Chen D (2011) SNR measurement based on linearity test for ADC BIST. In: Proceedings of the 2011 IEEE International Symposium of Circuits and Systems (ISCAS), pp 269–272. https://doi.org/10.1109/ISCAS.2011.5937553

[24] Duan J, Jin L, Chen D (2012) Testing ADC spectral performance without dedicated data acquisition. IEEE Trans Instrum Meas 61:2941–2952. https://doi.org/10.1109/TIM.2012.2202949

[25] Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13:21–27. https://doi.org/10.1109/TIT.1967.1053964

[26] Drucker H, Burges CJC, Kaufman L, Smola AJ, Vapnik V (1997) Support vector regression machines. In: Mozer MC, Jordan MI, Petsche T (eds) Advances in Neural Information Processing Systems 9. MIT Press, Cambridge, MA, pp 155–161.

[27] Cutler A, Cutler DR, Stevens JR (2012) Random forests. In: Zhang C, Ma Y (eds) Ensemble Machine Learning. Springer, New York, pp 157–175. https://doi.org/10.1007/978-1-4419-9326-7_5

[28] Friedman JH (2001) Greedy function approximation: A gradient boosting machine. Ann Stat 29:1189–1232. https://doi.org/10.1214/aos/1013203451

[29] Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp 785–794. https://doi.org/10.1145/2939672.2939785



**Rongjie Li** received the B.S. degree in automation from the Anhui University of Technology (AHUT), Maanshan, China, in 2019, and the M.S. degree in instrumentation engineering from the University of Electronic Science and Technology of China, Chengdu, in 2022. He is currently working toward the Ph.D. degree in instrumentation science and technology.His current research interests focus on mixed-signal integrated circuit modeling, calibration, and testing



**Houjun Wang** received his B.S. degree from Beihang University in 1982, his M.S. degree from the University of Electronic Science and Technology of China (UESTC) in 1985, and his Ph.D. degree from UESTC in 1991. He is currently a professor at the Shenzhen Institute for Advanced Study, UESTC. His current research interests focus on test signal processing, integrated circuit testing, and intelligent test technology and instruments.



**Jia Li** received the B.S. degree in Computer Science and Technology from Yunnan Normal University, Kunming, China, in 2007. She is currently working as a test engineer at Shence chips testing and certification., Ltd. Her current research interests include integrated circuit testing.

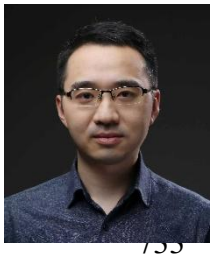


**Pu Pu** received the B.S. and M.S. degrees from the University of Electronic Science and Engineering of China (UESTC), Chengdu, China, in 2004 and 2014, respectively, where he is currently working toward the Ph.D. degree at the School of Automation Engineering. His research interests include integrated circuit test and mixed-signal chip test technologies in mass production testing, including high-speed and high precision AD converters, SoC, and RF chips.



**Guibing Zhu** received the B.S. and M.E. degrees in automation engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2011 and 2014, respectively, and the Ph.D. degree in advanced interdisciplinary studies from The University of Tokyo, Bunkyo, Japan, in 2019.

He is currently an Associate Professor at Shenzhen Institute for Advanced Study, UESTC. His research interests include wireless communications, embedded systems, and artificial intelligence.



**Wanyu Yang (Corresponding author)** obtained a bachelor's degree from the University of Electronic Science and Technology of China (UESTC) in 2005, and a master's degree from UESTC in 2008. He is currently a senior engineer at the Advanced Research Institute of UESTC (Shenzhen), with his main research focus on integrated circuit testing technology.