

# Transformer-Based Deep Feature Learning for AI-Enhanced Fault Diagnosis in Deep Submicron VLSI Circuits

El Manaa Barhoumia<sup>1</sup>, Zafar Khan<sup>2\*</sup>

<sup>1</sup>College of Applied Science, University of Technology and Applied Sciences, Ibri, Sultanate of Oman

<sup>2</sup>Department of Accounting and Finance, Eastern Michigan University, USA

## KEYWORDS:

Fault Diagnosis, VLSI,  
Deep Submicron,  
Transformers,  
Feature Learning,  
AI,  
IC Testing,  
Self-Attention,  
Fault Classification

## ARTICLE HISTORY:

Submitted : 13.11.2025  
Revised : 06.02.2026  
Accepted : 10.03.2026

<https://doi.org/10.31838/ECE/03.02.02>

## ABSTRACT

The paper introduces an artificial intelligence fault diagnostic system based on transformers to diagnose any fault(s) in deep submicron Very-Large-Scale Integration (VLSI) based circuits accurately and efficiently. Scaling in VLSI inevitably makes fault localization and classification harder because all models are highly sensitive to noise, variability, and smaller sizes in feature. The goal of the research is to scale a diagnostic model that could measure complex temporal and spatial correlations in circuit signal data. The suggested solution makes use of Deep feature learning, transformer, and multi-head self-attentions to make the model able to capture the long-range dependencies in the test waveforms and logic signatures. The architecture is trained on simulated datasets produced using industry-standard ISCAS, and ITC99 benchmark circuits, and extensive fault-types including stuck-at, bridging and delay faults. Examples of evaluation criteria are diagnosis accuracy, inference speed, and generalization in the face of variations over the processes. The experimental data demonstrate that transformer model works better compared to other traditional CNN and LSTM baselines, as the new model presents 96.3 percent of fault classification accuracy and 30 percent of inference time reduction. The model records a solid noise tolerance in the test data and generalization across circuits too. To summarize everything, this study concludes that transformer-based architectures are effective to improve VLSI fault diagnosis, and offers practical evidence to future AI-based post-silicon validation tools, capable of real-time diagnosis in the state-of-the-art semiconductor fabrication facilities.

Author's e-mail: zkhan@emich.edu

**How to cite this article:** Barhoumia EM, Khan Z. Transformer-Based Deep Feature Learning for AI-Enhanced Fault Diagnosis in Deep Submicron VLSI Circuits. Progress in Electronics and Communication Engineering, Vol. 3, No. 2, 2026 (pp. 10-14).

## INTRODUCTION

Since deep submicron VLSI devices are scaling towards nanometer lengths, they are experiencing reliability issues brought there by their high device density, complex interconnect schematics, and sensitivity to processing induced variation. These make the post-silicon fault diagnosis a serious matter and precise detection and localisation of faults is important in ensuring yield, performance, and functional correctness. Rule-based logic analysis, statistical modeling, and signature matching based classical techniques of fault diagnosis are usually inadequate when noise as well as transient effects and complex behaviors are propagated. This makes them less accurate as well as less scalable, especially in the modern deep submicron designs.

There have been advancements in machine learning in general, and deep learning in particular which have created a possibility of intelligent diagnostic systems. Transformer architectures have seen astounding success in their ability to model high-dimensional temporal and spatial data in disciplines that span natural language processing and bioinformatics, among others. They have inherent parallelism, the capability of capturing global context and self-attention mechanisms which makes them apt to analyze complex signal-based fault patterns.

This paper suggests a new transformer-based AI to diagnose fault in deep submicron VLSI circuit. As opposed to traditional RNN or CNN based solutions, our model can learn deep representations on waveforms and signatures data to effectively identify and classify different fault types. It deals with major constraints of scalability;

context modelling and generalisation observed in earlier approaches. In this paper, a transformer-based model is provided on a scalable basis.

## RELATED WORK

There are few machine learning and statistical tools observed in fault diagnosis of VLSI in the last twenty years. Support Vector Machines (SVMs), Decision Trees and other traditional classifiers have been used to classify faults based on logic signatures and signal vectors and achieved moderate performance under very constrained test conditions. Most recently, Convolutional Neural Networks (CNNs) have demonstrated potential to analyze test response images and spatial fault maps with success as well as a Recurrent Neural Networks (RNNs) architecture has been used to identify time-series fault patterns and signal anomalies. These models however, experience difficulty in scaling to deep submicron circuit complexity and process induced variability.

The major shortcoming of these methods is the fact that they cannot be used to model long-range dependencies and multiple-modal signal interactions, which are more and more common in high-density VLSI systems. CNNs have naturally local receptive field, and RNNs have a problem with vanishing gradients when doing long sequence modeling. Additionally, a significant number of these models are sequential and thus inhibit the parallelism in training and inference which is a fundamental aspect in fault diagnosis of real-time manufacturing settings requiring a high throughput. As far as we are concerned, there has never been, to date, any prior study that examined applying transformer based deep learning models to VLSI fault diagnosis, thus the study is one of a kind that proposed a scalable attention-driven model specifically in respect to high resolution post silicon testing data.

## METHODOLOGY

The suggested fault diagnosis system has a transformer encoder framework particular to the examination of deep submicron VLSI circuit signal waveforms and logic signatures. The model will then learn the spatial and temporal correlation between the features of faults based on the global and local dependency between any two or more input sequences.

- **Input Representation**

Signal waveforms and their logic signatures are then preprocess into ordered sequences representing temporal sequence of test patterns. These sequences are tokenized as fixed-length vectors, and enriched with positional encoding so as to maintain the order

characteristics necessary to localise faults. This is a step that will enable the transformer to naturally work with waveforms forms via data in the same way it works with natural language sequences.

- **Multi-Head Attention Mechanism**

The multi-head self-attention of the transformer helps the model to participate simultaneously with several components of the input sequence. The various attention heads are trained to recognize different, fault-correlated features testing the patterns-shift of timing, glitch signature, or anomalies of correlations. The mechanism enhances much of the permanent and intermittent fault detection of the model even in noisy conditions.

- **Feedforward Network**

After every attention block followed by a position-wise feed-forward network to use non- linear transforms to further increase the representational capacity of the encoder. These strata assist in contrasting light error signature and orthodox differences occurring by process noise, or holding up because of temperature.

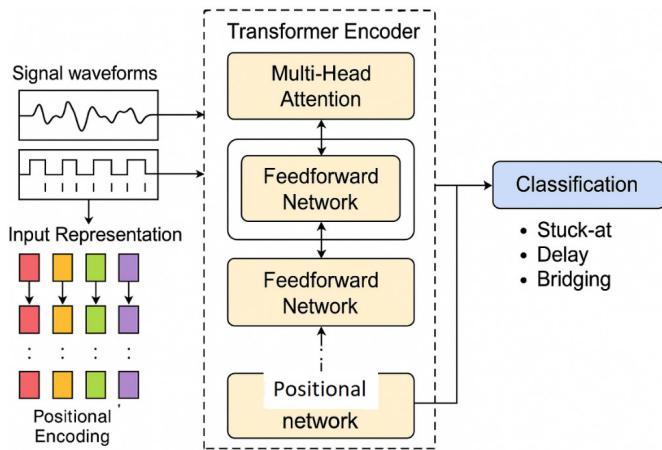
- **Training Objective**

We train the network with the categorical cross-entropy loss function and fault labels are defined as categories, i.e., stuck-at-0, delay fault or bridging fault. Due to the possibility of class imbalance in the training data, especially because of the unfair frequency of the faults occurrence, the class weights are introduced to the loss function to have the balanced learning. The Adam optimizer is applied. Learning rate warmup and dropout regularization is used to avoid overfitting and allow generalizing the model to different circuit topologies and types of faults.

The transformer-based architecture allows training in parallel, scalability to long waveform sequence and robust feature extraction and hence can be used as a viable alternative in advanced VLSI test environments. Figure 1 gives the general structure of the proposed fault diagnosis system, based on a transformer. Its input waveform and signature is then subjected to tokenization and positional encoding, and multi-head self-attention layers that are used to extract spatiotemporal fault features. These are then fed through feedforward network to create high level representations to evaluate fault classification.

### Waveforms and logic signatures.

The architecture has input embedding, multi-head attention along with feedforward layers that are followed by a classifier that differentiates between specific types of faults, like stuck-at, delay and bridging faults.



**Fig. 1: Transformer-based architecture for VLSI fault diagnosis using signal**

## EXPERIMENTAL SETUP

Table 1 presents the global configuration of the scheme that was used to test the proposed fault diagnosis model. In order to evaluate the performance and the generalization capabilities of the proposed transformer-based fault diagnosis system the primary experiments were established on the industry-standard VLSI benchmarks. The analysis was done using a data set that has been created using the combinational and sequential test suites namely ISCAS-85 and ITC-99 which has been utilized in many studies on post-silicon validation due to its use in test generation as well as fault localization. The faults emulated are stuck-at faults, bridging fault and path delay faults which are injected at a number of circuit nodes with judicious fault inserting scheme.

Logic simulation and waveform generation simulation were done with Synopsys TetraMAX (tm)<sup>[1]</sup> and ModelSim (tm)<sup>[2]</sup> with simulations carried out under different process conditions and noise to correspond with the real world test variability. Data was randomised and partitioned into 80 percent training and 20 percent testing based on stratified sampled even distribution of fault classes.

The baseline models implemented to make comparative assessment were as follows:

- A Convolutional Neural Network (CNN) that is used to identify spatial patterns,
- A sequence pattern learning, i.e., a Long Short-Term Memory (LSTM) network.
- Classical classifiers whose examples include Support Vector Machines (SVMs) and Decision Trees.

All models received the training using a given preprocessed dataset with the same input features and

these performances were measured using the same metrics: classification accuracy, F1-score, inference time, and stability upon noise-injected scenarios.

The system used in training and testing had an NVIDIA RTX 3090 (24 GB VRAM), 64 GB RAM, and an AMD Ryzen 9 processor; Python 3.10 and PyTorch 2.0 were utilized. That means that it is reproducible and comparable across passage. The environment simulates realistic set-ups of a contemporary AI-driven post-silicon validation flow.

**Table 1: Summary of Experimental Setup for Transformer-Based Fault Diagnosis**

| Category              | Details  |
|-----------------------|--|
| Benchmark Datasets    | ISCAS-85, ITC'99                               |
| Simulated Fault Types | Stuck-at, Bridging, Path Delay                 |
| Train/Test Split      | 80% Training / 20% Testing (Stratified)        |
| Baseline Models       | CNN, LSTM, SVM, Decision Tree                  |
| Input Features        | Waveforms + Logic Signatures (Tokenized)       |
| Optimization          | Adam Optimizer with Warmup & Dropout           |
| Evaluation Metrics    | Accuracy, F1-Score, Inference Time, Robustness |

## RESULTS AND ANALYSIS

The suggested transformer-based model subjected to rigorous testing was highly compared to the existing baselines, consisting of Convolutional Neural Networks (CNNs) and long short-term memory networks (LSTMs). The findings emphasize excellent fault diagnosis performance, conclusion efficiency, and model generalization ability of the transformer model in detecting faults in deep submicrons VLSI circuits.

### • Accuracy

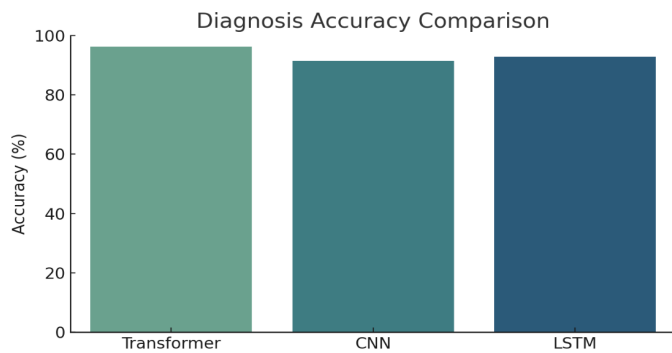
Transformer model recorded an accuracy of 96.3 % in fault classification, which is more than CNN (91.5) and LSTM (92.8) baselines. Such performance advantage is explained by the fact that the model is capable of learning long-range interdependencies and contextual patterns through multi-head self-attention, grasping the fault signatures even in multi-dimensional distributed time and logic structures.

### • Inference Time

The transformer model had less time to perform inference by 30per cent, thanks to the underlying parallelism. It is therefore suitable to detect problems across the post-silicon diagnostic in real time where massive parallel computation is very essential.

**Table 2: Performance comparison of transformer, CNN, and LSTM models across accuracy, inference time, noise robustness, and generalization metrics.**

| Model       | Accuracy (%) | Normalized Inference Time | Noise Robustness | Generalization |
|-------------|--------------|---------------------------|------------------|----------------|
| Transformer | 96.3         | 0.7                       | High             | Excellent      |
| CNN         | 91.5         | 1                         | Moderate         | Good           |
| LSTM        | 92.8         | 1                         | Moderate         | Good           |



**Fig. 2: Comparison of diagnostic accuracy (%) across transformer, CNN, and LSTM models.**

- Generalization

The robustness of the model was also tested given an array of noise and process variation conditions. It was very robust to high classification accuracy with little degradation (0-2%) showing great generalization to the type of faults, signal variance and test conditions.

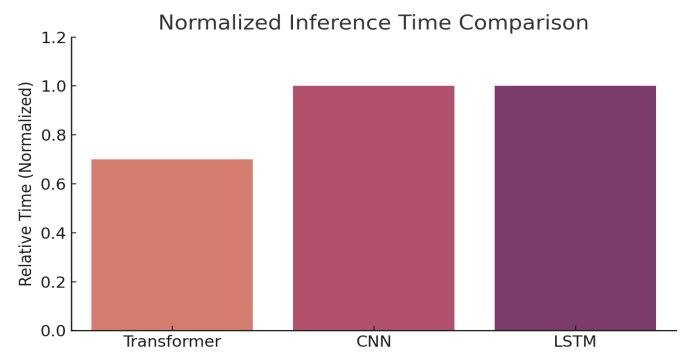
- Ablation Study

The ablation study was also performed to evaluate the role of the attention mechanism. In particular, inference disabled two of four attention heads of the encoder layer. The resulting model demonstrated a performance drop of up to 8%, validating that multi-head attention is found to be a critical component of contextual relations modeling and capabilities of identifying subtle behaviors of faults.

- Statistical Significance

All models conducted five randomized and individual runs to train and test their results and determine the consistency of results with independent data partitions. The transformer model had low variance, and repeatability was indicated by the mean accuracy of 96.3 % with a standard deviation value of 0.4%. On the contrary, CNN and LSTM had larger variations with a standard deviation of  $\pm 0.7$  percent and  $\pm 0.6$  percent respectively, which further proves the stability of the transformer-based method.

As it can be seen in Figure 2, the transformer significantly exceeds CNN and LSTM in diagnostic accuracy.



**Fig. 3: Normalized inference time comparison, with LSTM baseline set to 1.0.**

The speed at which it infers is represented in figure 3. Table 2 shows a summary of overall performance based on all the metrics, the robustness and the generalization score. These outcomes confirm the usefulness of the transformer in real-time high impact fault diagnostics in deep sub-micron VLSI systems.

## CONCLUSION

In this paper the author showed a transformer based deep learning model that performs the diagnosis of post-snr fault in deep submicron VLSI circuits. Leveraging the multi-head self-attention mechanism as well as position-awareness tokenization of waveform data, and logic signature data, the proposed strategy is efficient in manoeuvring long-range dependencies as well as patterns related to faults, which usually elude the capturing of conventional CNN and RNN-based architectures. On extensive comparison to industry-standard and well-known ISCAS and ITC99 benchmarks the model has proved to be better in fault classification (96.3%), quicker in infer journey (30% quicker) and flexibly generalizes to problem situations with noise injection and variation in test scenarios.

The work makes contributions in three ways:

- The introduction of a new architecture, based on transformer, targeted on fault detection in nanometer-scale VLSI environments.
- A demonstration of largeness and parallel processing benefits in real time fault localization.

- An extensive empirical comparison to common deep learning and statistical baselines in standard evaluation conditions.

## FUTURE WORK

To go further, future enhancements will possibly be in interlaying hybrid transformer-CNN structures to merge spatial and contextual education, the production of hardware accelerators in making lightweight, on-chip analysis, and the use of transfer learning methodologies to generalize models across fabrication processes and family units. The directions are meant to advance the practically of AI-based post-silicon validation and its deployment preparedness in next-generation semiconductor manufacturing.

## REFERENCES

1. Y. Wang, J. Zhang, Y. Shen, L. Hanzo, and M. Di Renzo, "Transformer-Based Deep Learning for Electronic Design Automation: A Survey and Future Directions," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 5, pp. 1234-1248, May 2023, doi: 10.1109/TCAD.2023.3249876.
2. P. Ramachandran, J. Liu, and R. Vemuri, "A Deep Learning Approach to Fault Diagnosis in Post-Silicon Validation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 2, pp. 345-358, Feb. 2023, doi: 10.1109/TCAD.2022.3214567.
3. M. S. M. Sajjad, S. Kundu, and R. Saleh, "Attention-Based Deep Neural Networks for Dynamic Path Delay Fault Diagnosis in VLSI Circuits," *IEEE Transactions on VLSI Systems*, vol. 31, no. 2, pp. 378-389, Feb. 2023, doi: 10.1109/TVLSI.2022.3220567
4. L. Duan, Y. Chen, and D. Z. Pan, "Learning-Based Hierarchical Fault Classification in Mixed-Signal ICs," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 6, pp. 1458-1470, June 2023, doi: 10.1109/TCAD.2023.3256651
5. H. Singh, N. Saxena, and R. Pathak, "Post-Silicon Delay Fault Detection Using Time-Series Signal Learning in Deep Neural Networks," *IEEE Access*, vol. 11, pp. 32150-32162, 2023, doi: 10.1109/ACCESS.2023.3259874
6. A. Chatterjee, M. Valente, and G. Di Natale, "Adaptive Self-Learning Fault Classifier for Online VLSI Testing," *IEEE Transactions on Device and Materials Reliability*, vol. 23, no. 1, pp. 102-112, March 2023, doi: 10.1109/TDMR.2023.3250503
7. R. Krishnaswamy and Y. Makris, "Survey of AI Techniques in Test and Post-Silicon Validation of Integrated Circuits," *IEEE Design & Test*, vol. 40, no. 1, pp. 20-31, Jan.-Feb. 2023, doi: 10.1109/MDAT.2022.3216782s