

DeepFeatIoT: Unifying Deep Learned, Randomized, and LLM Features for Enhanced IoT Time Series Sensor Data Classification in Smart Industries

Muhammad Sakib Khan Inan and Kewen Liao*

School of Information Technology, Deakin University, Australia
sakib.inan@research.deakin.edu.au, kewen.liao@deakin.edu.au

Abstract

Internet of Things (IoT) sensors are ubiquitous technologies deployed across smart cities, industrial sites, and healthcare systems. They continuously generate time series data that enable advanced analytics and automation in industries. However, challenges such as the loss or ambiguity of sensor metadata, heterogeneity in data sources, varying sampling frequencies, inconsistent units of measurement, and irregular timestamps make raw IoT time series data difficult to interpret, undermining the effectiveness of smart systems. To address these challenges, we propose a novel deep learning model, DeepFeatIoT, which integrates learned local and global features with non-learned randomized convolutional kernel-based features and features from large language models (LLMs). This straightforward yet unique fusion of diverse learned and non-learned features significantly enhances IoT time series sensor data classification, even in scenarios with limited labeled data. Our model's effectiveness is demonstrated through its consistent and generalized performance across multiple real-world IoT sensor datasets from diverse critical application domains, outperforming state-of-the-art benchmark models. These results highlight DeepFeatIoT's potential to drive significant advancements in IoT analytics and support the development of next-generation smart systems.

1 Introduction

In the era of Industry 4.0, the world around us is surrounded by a vast number of IoT sensors capturing data from diverse real-world environments, enabling smart automated actions or decision-making across various critical sectors, including smart cities, smart buildings, smart agriculture, and smart healthcare. These sensors record a wide range of real-world observations such as temperature, humidity, traffic flow, snow height, visibility, and more [Montori *et al.*, 2023; Georgakopoulos *et al.*, 2020]. A recent report from DemandSage¹ predicts that the number of connected IoT de-

vices will reach 40 billion by 2030. The vast amount of time series sensor data generated from these devices fuels essential smart industrial applications, such as automated traffic monitoring in cities, environmental monitoring for sustainability, smart building management, airport operations, and precision agriculture. These advancements are driven by the advanced automatic analysis of historical time series sensor data [Wajgi *et al.*, 2023; Manoharan *et al.*, 2021; Rajak and Selot, 2021].

With the growing number of connected IoT devices, metadata loss, which refers to the supporting textual data that describes sensors and their observations, has become a common issue [Čulić Gambiroža *et al.*, 2023]. This can result from network communication failures, battery depletion, a lack of standardized data storage and sharing systems [Čulić Gambiroža *et al.*, 2023; Montori *et al.*, 2023; Inan *et al.*, 2023], or, in some cases, metadata may not be stored at all due to security concerns [Stoyanova *et al.*, 2020]. The absence of accurate metadata poses significant challenges for researchers and developers, rendering large volumes of historical IoT time series sensor data uninterpretable [Čulić Gambiroža *et al.*, 2023; Inan *et al.*, 2023]. This, in turn, undermines automated data analysis and smart decision-making processes, which are essential for a wide range of critical industrial applications. In such cases, we are left with only time series sensor readings from IoT devices. Before conducting any data analysis, we must first determine the specific type of IoT sensor (*e.g.*, temperature, humidity, traffic flow) to which each time series data stream corresponds. This necessity gives rise to the challenge of IoT time series sensor data classification. Furthermore, classifying IoT time series sensor data is highly challenging due to its inherent complexity, which arises from sensors generating data from heterogeneous sources, various geographic locations, different sampling frequencies, and irregular timestamps [Montori *et al.*, 2023; Čulić Gambiroža *et al.*, 2023]. Additionally, manual, human-involved IoT data annotation or classification is time-consuming, labor-intensive, and often financially infeasible due to its high costs [Montori *et al.*, 2023].

To address these issues, several artificial intelligence (AI) researchers have developed state-of-the-art machine learning [Montori *et al.*, 2023; Čulić Gambiroža *et al.*, 2023] and deep learning algorithms [Inan *et al.*, 2023] for the effective classification of IoT time series sensor data. Previously, an ensem-

*Corresponding Author

¹<https://www.demandsage.com/number-of-iot-devices/>

ble of multiple machine learning methods, using improved model selection and class filtering strategies, was found to outperform traditional time series classification methods for IoT sensor data [Montori *et al.*, 2023]. Although ensemble machine learning showed promising performance, it struggled to fully capture the heterogeneity in time series patterns (a mixture of local sub-patterns and global patterns) due to its reliance on pre-computed statistical properties from IoT time series data. More recently, deep learning algorithms [Inan *et al.*, 2023] have been introduced to learn complex local and global features directly from raw IoT time series data to facilitate classification. However, these models still face challenges in generalizing well when labeled data is limited in heterogeneous IoT sensor datasets. This limitation arises because deep learning algorithms typically require large amounts of training data, which is not always readily available in IoT sensor data classification.

Inspired by the recent success of large language models (LLMs) like GPT (Generative Pretrained Transformer) in Natural Language Processing [Min *et al.*, 2023] and Computer Vision [Khan *et al.*, 2022], as well as the effectiveness of randomized convolutions in capturing diverse time series patterns [Dempster *et al.*, 2020; Tan *et al.*, 2022], we propose a novel deep learning model for IoT time series sensor classification. Our approach unifies learned multi-scale convolutional and recurrent representations with non-learned randomized convolutional features and pre-trained LLM features to enhance model generalization across heterogeneous IoT sensor datasets. The main contributions of this work are summarized as follows:

- We propose a novel deep learning model, DeepFeatIoT, which uniquely integrates learned, randomized, and pre-trained LLM features to enhance classification performance across various IoT sensor datasets, even in scenarios with limited labeled data.
- We employ a uniform dense feature transformation module to fuse feature representations of different scales, ensuring balanced feature contributions. Through this module, our randomized convolutional features and pre-trained LLM features (*e.g.*, from GPT-2 [Radford *et al.*, 2019]) are effectively incorporated.
- We establish a comprehensive benchmark through rigorous empirical evaluations against state-of-the-art deep learning models across multiple real-world IoT sensor datasets, highlighting the importance of feature diversity for improved classification.

To the best of our knowledge, we are the first to effectively fuse a range of learned and non-learned features within a deep learning model for classifying heterogeneous IoT sensor types, supporting critical applications in smart industries.

2 Related Works

This section is organized into two subsections: *IoT Time Series Sensor Data Classification* and *Fundamental Time Series Classification*. The first subsection reviews state-of-the-art methods specifically designed for IoT time series sensor data,

while the second highlights recent algorithmic advancements in general time series classification.

2.1 IoT Time Series Sensor Data Classification

IoT time series sensor data are often noisy, with highly correlated patterns across class labels, overlapping sub-patterns, and inherent heterogeneity due to variations in timestamp ranges, sampling ratios, frequencies, or units of measurement. This makes the classification of IoT sensors from time series data a challenging task. Additionally, irregularities in time series intervals, unlike traditional time series classification problems where data is equally spaced in time, further complicate the process. Due to the popularity of statistical transformation and data mining methods in the time series classification domain, researchers previously developed a probabilistic data mining approach [Calbimonte *et al.*, 2012] that incorporated slope distribution computation via linear approximation of IoT time series sequences for heterogeneous IoT sensor data classification. Postol *et al.* [Postol *et al.*, 2019] later proposed a topological data analysis-based strategy for noisy IoT sensor data classification, while Borges *et al.* [Borges *et al.*, 2022] introduced a transformation-based classification strategy that converts raw sensor data into ordinal patterns, improving feature representation and class separability. However, statistical transformation methods often struggle due to their assumptions about data distribution, especially with the variability and non-linearity of time series patterns, which may involve a mix of local and global patterns. Although these methods may work well for specific datasets, recent studies have found them to be ineffective across a variety of heterogeneous IoT sensor data classification scenarios [Montori *et al.*, 2023]. Recently, ensemble machine learning has outperformed statistical transformation and data mining methods, offering better generalizability, as demonstrated by Montori *et al.* [Montori *et al.*, 2023] and Gambiroza *et al.* [Čulić Gambiroža *et al.*, 2023]. However, despite their success, ensemble machine learning methods often suffer from overfitting in many use cases. Very recently, deep learning has emerged as a promising approach, leveraging the ability to learn global and local patterns in parallel to improve classification performance [Inan *et al.*, 2023]. Deep learning algorithms, however, require specific pre-processing or augmentations in cases where the number of samples is limited, particularly when there is class imbalance. This could potentially lead to adding bias or cause data-leakage.

2.2 Fundamental Time Series Classification

In fundamental time series classification problems, data points are typically equally spaced over time, and the dataset generally contains time series from the same or similar domains or applications, which is often different in the case of IoT time series [Bagnall *et al.*, 2017; Čulić Gambiroža *et al.*, 2023]. Over the years, fundamental time series classification methods have seen significant growth and advancement. It began with classical distance-based methods like Dynamic Time Warping (DTW) [Bagnall *et al.*, 2017] and has evolved to more recent methods, such as the Large Language Model-based approach GPT4TS [Zhou *et al.*, 2024]. Deep learning in time series classification gained attention when

Ismail et al. [Ismail Fawaz *et al.*, 2020] introduced the state-of-the-art convolutional architecture InceptionTime, inspired by AlexNet [Krizhevsky *et al.*, 2012]. Recently, randomized convolutions became popular in the time series classification space with the introduction of scalable random convolutional kernel-based methods like ROCKET [Dempster *et al.*, 2020] and its extension MultiROCKET [Tan *et al.*, 2022]. Self-attention-based transformer architectures have also shown promising performance, particularly with the introduction of the patching concept for time series, proving the effectiveness of dividing a time series sequence into 64 patches for use in the transformer architecture named PatchTST [Nie *et al.*, 2022]. Later, the researchers of PatchTSMixer [Ekambaram *et al.*, 2023] developed a Multi-Layer Perceptron (MLP)-based approach using the patching concept over time series sequences, making it more lightweight and faster than previous self-attention-dominant methods like PatchTST [Nie *et al.*, 2022]. The growing popularity of large language models (LLM) has also inspired AI researchers in the time series domain to develop a Generative Pre-trained Transformer (GPT)-based method, GPT4TS [Zhou *et al.*, 2024], which has achieved competitive performance in fundamental time series tasks.

While these methods perform well in fundamental time series classification problems, due to the unique nature of IoT time series data, similar methods have been found to struggle or be ineffective in a wide range of heterogeneous IoT time series classification problems [Inan *et al.*, 2023; Montori *et al.*, 2023]. Although there have been recent developments in deep learning algorithms specifically designed for IoT time series data classification, as previously discussed, they still suffer from overfitting when faced with a limited number of labelled samples along with class imbalance [Inan *et al.*, 2023]. Thus, a research gap remains in advancing state-of-the-art methods for improved feature extraction to support the classification of IoT time series sensor data, despite the limited labelled data challenges present in the data.

3 Method: DeepFeatIoT

In this section, we delineate the details of our proposed deep learning model, DeepFeatIoT. An overview of DeepFeatIoT is illustrated in Figure 1. DeepFeatIoT takes raw IoT time series sensor data as input (without any preprocessing) and extracts deep-learned local and global features, randomized convolutional features, and pre-trained LLM features. These four diverse sets of latent feature representations are then combined through an optimized dense feature transformation (DFT) module, leading to improved classification of IoT time series sensor data. The following subsections provide a detailed methodology of the proposed DeepFeatIoT model.

3.1 Preliminaries

An individual sample of IoT sensor data from the whole dataset is essentially a vector containing a sequence of real numbers, denoted as $X_i = (x_1, x_2, x_3, \dots, x_n)$, with i indexing the i -th sample from the dataset and n representing the number of timestamps. Each x_j denotes a numerical value of sensor reading at the j -th timestamp. Then the task of IoT

sensor data classification can be formulated as determining the value of y_i , which is the designated class label (*e.g.*, temperature, humidity, light sensor, etc.) for the i -th sample X_i (time series sequence) from the dataset.

3.2 Learned Local and Global Features

As IoT sensor data generates from heterogeneous sources and the datasets contain sensors from diverse domains, it is important to extract both global pattern and local sub-pattern from IoT time series for better feature representation. In this regard, we adopted a learned feature extraction strategy proposed in a previous study [Inan *et al.*, 2023] with some minor modification. In our proposed model, we incorporated a stack of bi-directional gated recurrent units (Bi-GRU) layers to extract deep learned global features and a stack of learned non-dilated convolutional (conv) kernels (conv layers with different kernel sizes) to extract deep learned local features. Moreover, in the learned global feature extraction process, to mitigate the vanishing gradient problem in recurrent neural networks [Liu *et al.*, 2021], we applied the ReLU (Rectified Linear Unit) transformation to clip negative values in the global learned latent feature representation. This learned feature extraction process can be mathematically represented as,

$$\mathcal{F}_g = \sigma(G(X_i)), \quad (1)$$

where X_i is the raw IoT time series sequence as describe earlier, G is the function for computing global feature using Bi-GRU layers and σ is the ReLU activation, then the resultant is global learned feature vector \mathcal{F}_g . On the other hand, features from multi-scale convolutional blocks is concatenated along the depth dimension, which can be similarly represented mathematically as:

$$\mathcal{F}_c = \mathcal{K}_3(X_i) || \mathcal{K}_5(X_i) || \mathcal{K}_7(X_i) || \mathcal{K}_{11}(X_i), \quad (2)$$

where \mathcal{K}_i represents function with stack of convolutional layers of kernel size i and \mathcal{F}_c represents latent feature vector comprising of deep learned local features.

3.3 Randomized Features

Randomized non-learned convolutional kernels based feature transform of input time series were proven to effective for capturing discriminating features for fundamental time series classification by several previous studies [Dempster *et al.*, 2020; Jiménez and Raj, 2019]. Large amount of randomized convolutional kernel transformation can produce variety of diverse non-linear features that could be effectively used for further classification task. This method is scalable and works well for small time series datasets with limited sample and labels, effectively avoiding overfitting due to its unsupervised nature in feature extraction stage [Dempster *et al.*, 2020]. In this regard, considering the randomness in temporal pattern for IoT time series sensor data due to inherent heterogeneity, noise and data sparsity issue, randomized convolutional kernels could be a scalable solution to capture diverse features with different aspects. So, we incorporate ten thousand randomized convolutional kernel over raw IoT time series sequence (X_i) inspired by the works of [Dempster *et al.*, 2020] with modified and improved parameter configurations. In our proposed model, for extraction of randomized

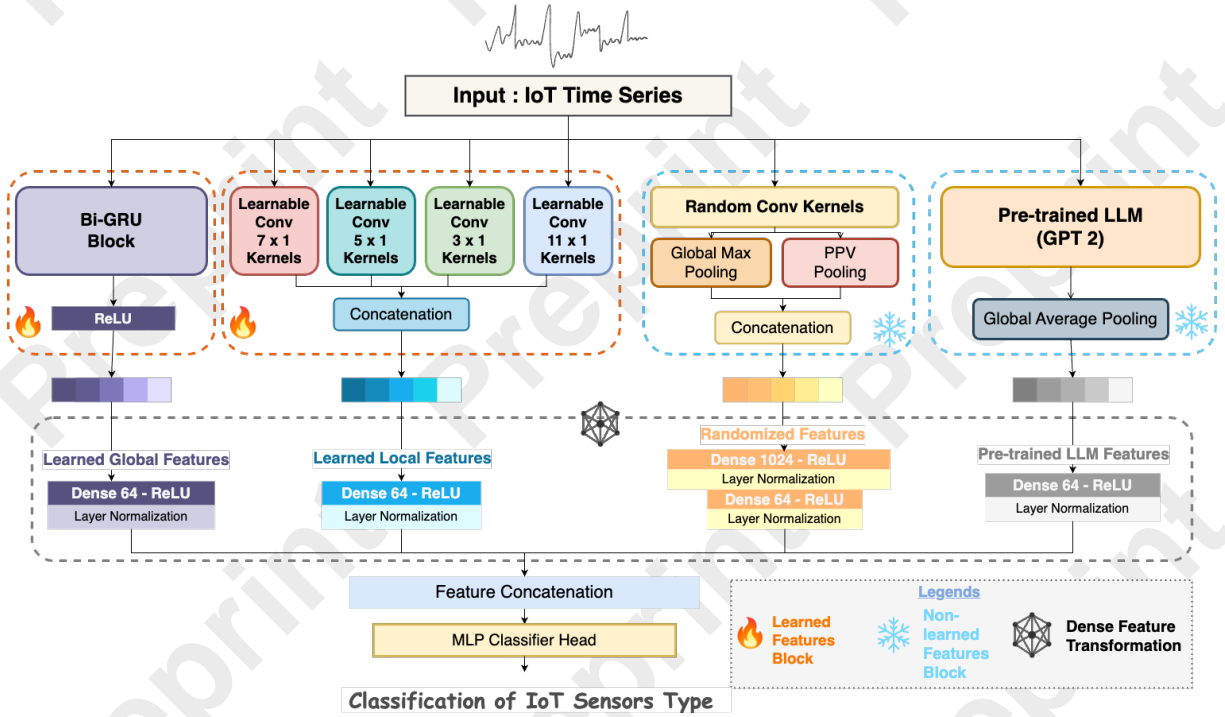


Figure 1: Illustration of the proposed deep learning model DeepFeatIoT.

features, the kernel weights (w_i) of the randomized kernels (C_i) were drawn from the normal distribution of $\mathcal{N}(0, 0.05)$. The length (k_i) of all kernels was set to 9. Unlike from previous work [Dempster *et al.*, 2020], we set the bias (b_i) term to zero and use a fixed dilation (d_i) size 4. These deterministic modifications of parameter values were made to make the model more generalized across IoT time series sensor data. This process can be represented as:

$$\mathcal{F}_r = \sum_{i=1}^m C_i \{w_i, k_i, b_i, d_i\}(X_i), \quad (3)$$

where \mathcal{F}_r represents latent feature vector of randomized convolution kernels (the number of kernels $m = 10000$) aggregated via global max pooling and proportion of positive value pooling. The updated parameter configuration is also more scalable due to its fixed dilation and padding, while the statistical distribution of weights has been reduced to a standard deviation of 0.05.

3.4 Pre-trained LLM Features

Pre-trained LLMs designed for solving natural language processing tasks, have already proven effective at capturing long-term sequential patterns from complex data and can be used to solve a variety of challenging problems [Yang *et al.*, 2024]. Especially, when there is limited availability of data and labels in downstream tasks. Time series data has a similar structure to textual data [Gruver *et al.*, 2024], as both text and time series data exhibit sequential nature and correlation. The main difference is that, in time series data, the elements of a sequence are real-valued numbers, whereas in

the text domain it is generally sequence of words in natural language [Gruver *et al.*, 2024]. Inspired by the works of [Zhou *et al.*, 2024], to utilize the benefit of pre-trained LLM features in the context of IoT time series sensor data, we incorporate pre-trained GPT2 [Radford *et al.*, 2019] in our proposed model for extracting LLM based sequential contextual features from raw IoT time series sequence. Unlike the previous work [Zhou *et al.*, 2024], in the proposed model, we do not do any re-programming or transformation of raw IoT time series, instead, a IoT time series sequence X_i , is tokenized as textual sentence (numbers treated as sequence of characters or words) where each element or value, x_j at individual timestamp is separated by special characters to make up a text sequence and fed as input to the GPT2 [Radford *et al.*, 2019] model after tokenization. We incorporate all the 12 layers (transformer blocks) of GPT2 in its original form. This would help the model to take benefit of pre-training on text data and extract sequential correlation in IoT time series without any transformation or re-programming. This can be mathematically represented as

$$\mathcal{F}_l = \text{pool}_{avg}(\text{GPT2}(\{X_i\})). \quad (4)$$

Here, a two-dimensional feature vector is generated as output by the GPT2 model which goes through a Global Average Pooling layer (pool_{avg}) to produce one-dimensional latent features space vector \mathcal{F}_l , that represents our proposed pre-trained LLM features capturing sequential contextual correlation in raw IoT time series data.

3.5 Dense Feature Transformation

After extracting the learned latent feature vectors \mathcal{F}_g and \mathcal{F}_c , along with the randomized features \mathcal{F}_r and pre-trained LLM

features \mathcal{F}_l , the next step is to optimally combine these four diverse features into a single latent space to facilitate the classification step. These feature vectors vary in dimensionality or feature space size. For instance, the dimensionality of \mathcal{F}_r and \mathcal{F}_l are $\mathcal{F}_r \in \mathbb{R}^{20000 \times 1}$ and $\mathcal{F}_l \in \mathbb{R}^{768 \times 1}$, which are much larger compared to the dimensionalities of \mathcal{F}_g and \mathcal{F}_c , which are $\mathcal{F}_g \in \mathbb{R}^{128 \times 1}$ and $\mathcal{F}_c \in \mathbb{R}^{256 \times 1}$, respectively.

At this stage, directly concatenating these feature vectors with varying dimensions as input to the next fully connected feed-forward layer (FCN) would result in bias where the larger and sparser feature vectors would dominate over smaller feature vectors in terms of dimensions. This could lead to overfitting, poor generalization, and suffering from the curse of dimensionality. To mitigate these issues, we designed the Dense Feature Transformation (DFT) module, a simple yet effective method for transforming each feature vector into a dense vector space with reduced dimensionality. This acts as an indirect step for the selection and scaling of features within the neural network. The DFT module transforms the four different feature vectors into separate dense vector spaces of equal size (64 dimensions), addressing the issues mentioned above. It can be represented in simplified mathematical form as:

$$\mathcal{F}_d = \sigma(W_g\{\mathcal{F}_g\}) \parallel \sigma(W_c\{\mathcal{F}_c\}) \parallel \sigma(W_r\{\mathcal{F}_r\}) \parallel \sigma(W_l\{\mathcal{F}_l\}), \quad (5)$$

where σ represents the non-linear activation function, which in this case is ReLU, and W_g, W_c, W_r, W_l are the weight matrices applied respectively to $\mathcal{F}_g, \mathcal{F}_c, \mathcal{F}_r, \mathcal{F}_l$. Among these, W_g, W_c , and W_l are dense layers with a size of \mathbb{R}^{64} , while W_r consists of a stack of two dense layers with sizes \mathbb{R}^{1024} and \mathbb{R}^{64} , respectively. Each dense fully connected layer is followed by a layer normalization operation. The output of DFT module is \mathcal{F}_d a combined feature vector containing all 4 types features represented in a single latent space. After that, as illustrated in Figure 1, the combined feature vector is fed into an MLP head, which consists of a stack of two feed-forward neural network layers with 128 and 64 neurons, respectively. Each layer is followed by layer normalization and a dropout of 50%. Finally, the output from the last hidden layer is passed through a softmax layer for the classification task.. Algorithm 1 presents the pseudo-code of the proposed DeepFeatIoT algorithm, aligning with the brief description of the individual modules outlined above. The proposed DeepFeatIoT model incorporates a wide range of learned and non-learned features to capture diverse contextual aspects from raw IoT time series, enhancing generalizability across various IoT sensor data domains, even in scenarios with limited data and label availability, without requiring any data pre-processing. The DeepFeatIoT model is trained for 200 epochs, and the best weights, based on testing accuracy scores, are used for the final experimental evaluation. The training is conducted using the popular Adam optimizer with an initial learning rate of 0.001, combined with inverse time decay every 100 steps, and categorical focal cross-entropy loss.

4 Experimental Results

This section discusses the experimental designs and validations of our proposed model across IoT sensor datasets con-

Algorithm 1 DeepFeatIoT Algorithmic Pseudo-code

Input: X_i represents a particular IoT time series sequence from the IoT sensor dataset, which can be mapped to a particular class label y_i . So, a IoT sensor dataset can be represented as $\mathcal{D} = \{(X_i, y_i), \dots, (X_n, y_n)\}$. For a classification problem, the model is going to be trained iterating over every i th instance in the dataset, \mathcal{D} (n represents number of samples/instances in the dataset).

- 1: **for** Each *epoch* in *Epochs* **do**
- 2: For each i th instance (X_i) compute learned features \mathcal{F}_g and \mathcal{F}_c using accordingly equations (1) and (2).
- 3: Similarly, compute randomized convolutional kernel feature \mathcal{F}_r using equation (3) and extract pre-trained LLM features, \mathcal{F}_l , from LLM using equation (4).
- 4: Combine features $\mathcal{F}_g, \mathcal{F}_c, \mathcal{F}_r, \mathcal{F}_l$ into a single feature vector space \mathcal{F}_d using (5).
- 5: \mathcal{F}_d goes into MLP Head and through softmax layer and generate probability distribution to determine the value of y_i .
- 6: Compute errors in prediction after each batch and update weights.
- 7: **end for**

Output: A trained deep learning model

taining time series sensor data serving various critical applications and domains. It also presents a benchmark analysis against previous state-of-the-art studies. Each IoT sensor dataset was resampled (stratified) into a training and testing ratio of 70:30 (random seed 100) [Montori *et al.*, 2023]. The detailed programming or source code implementation of this research can be found at the following GitHub repository: <https://github.com/skinan/DeepFeatIoT-IJCAI-2025>

Furthermore, to establish robust benchmark performance comparison baselines, we include state-of-the-art IoT sensor classification models such as MACE [Montori *et al.*, 2023] and DeepHeteroIoT [Inan *et al.*, 2023], as well as advanced time series classification models, including convolution-based InceptionTime [Ismail Fawaz *et al.*, 2020], and H-InceptionTime [Ismail-Fawaz *et al.*, 2022], transformer-based PatchTST [Nie *et al.*, 2022], MLP-mixer-based PatchTSMixer [Ekambaram *et al.*, 2023], LLM-based GPT4TS [Zhou *et al.*, 2024], randomized convolution based ROCKET [Dempster *et al.*, 2020] and MultiROCKET [Tan *et al.*, 2022] in the evaluation stage. As evaluation metrics, we consider two popular and standard evaluation metrics Accuracy and macro-average F1 score. For the fair comparison of performance evaluation, all the deep learning models are

Dataset	Length	Samples	Classes	Domain
Swiss	445	346	11	Smart City
Urban	864	1065	16	Smart City
Iowa	168	1000	8	Smart Airport
SBAS	168	255	5	Smart Buildings

Table 1: A summary of IoT time series datasets and respective critical industrial application domains

	Swiss		Urban		Iowa		SBAS		Average	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
InceptionTime	67.31	58.93	90.31	80.76	<u>96.67</u>	<u>96.70</u>	97.40	97.42	87.92	83.45
H-InceptionTime	75.00	62.66	90.94	81.01	<u>96.67</u>	<u>96.70</u>	97.40	97.41	90.00	84.45
PatchTST	62.50	56.86	85.63	74.54	87.99	87.96	89.62	89.50	81.44	77.22
PatchTSMixer	71.15	57.41	90.31	84.80	92.00	92.02	100.00	100.00	88.37	83.56
GPT4TS	70.19	62.54	86.56	80.98	88.33	88.22	100.00	100.00	86.27	82.94
ROCKET	75.00	72.13	90.31	82.81	91.00	91.02	97.40	97.48	88.43	85.86
MultiROCKET	84.62	82.76	96.56	95.23	95.00	94.96	<u>98.70</u>	<u>98.74</u>	93.72	92.92
MACE	86.54	<u>84.01</u>	92.25	88.74	91.01	91.04	97.40	97.49	91.80	90.32
DeepHeteroIoT	<u>86.54</u>	82.92	<u>96.88</u>	94.44	96.01	96.05	100.00	100.00	<u>94.86</u>	<u>93.35</u>
DeepFeatIoT	91.35	90.98	97.19	<u>94.79</u>	99.33	99.33	100.00	100.00	96.97	96.28

Table 2: Comparison of classification accuracy (acc) and macro-average F1 (F1) score of proposed model against state-of-the-art models across IoT Sensor datasets. **Bold**: best results, Underline: second best results

trained for 200 epochs with a learning rate of 0.001.

The IoT sensor datasets utilized in this study are: Swiss Experiment (Swiss) [Montori *et al.*, 2023], Urban Observatory² (Urban) [Montori *et al.*, 2023], Iowa ASOS³ (Iowa) [Inan *et al.*, 2023], Smart Building Automation System (SBAS) [Hong *et al.*, 2017]. Among all these datasets, the Swiss dataset contains a high level of noise and heterogeneity along with the issue of class imbalance and a very limited number of labelled samples [Calbimonte *et al.*, 2012; Montori *et al.*, 2023]. The Urban dataset contains highly correlated IoT sensor data.

On the other hand, Iowa and SBAS datasets are balanced in terms of number of samples per class. Table 1 summarizes key information about sensor data sets, including the length of time series, the number of samples, the number of classes, and the domain of critical industrial applications.

4.1 Classification Performance

Table 2 summarizes the classification Accuracy (Acc) and the corresponding macro-average F1 (F1) scores of our proposed model, DeepFeatIoT, compared with several state-of-the-art baselines evaluated on diverse IoT sensor datasets. The table reports the best results obtained from 10 independent runs for each model. Notably, DeepFeatIoT achieves a maximum accuracy of 96.97% and an average F1 score of 96.28%, thereby surpassing the second-best model, DeepHeteroIoT [Inan *et al.*, 2023], by more than 2%. Overall, when considering both metrics, the top five ranked models are: DeepFeatIoT (proposed), DeepHeteroIoT [Inan *et al.*, 2023], MultiROCKET [Tan *et al.*, 2022], MACE [Montori *et al.*, 2023], and H-InceptionTime [Ismail-Fawaz *et al.*, 2022], with average accuracies close to 90% or higher.

DeepFeatIoT consistently outperforms all benchmark models in accuracy on an individual dataset basis. For instance, on the heterogeneous Swiss dataset, DeepFeatIoT attains an F1 score of 90.98, which is approximately 6% higher than that of MACE [Montori *et al.*, 2023] and about 8% higher than those of both MultiROCKET [Tan *et al.*, 2022] and DeepHeteroIoT [Inan *et al.*, 2023]. In the Urban

dataset, although MultiROCKET [Tan *et al.*, 2022] achieves a marginally higher F1 score, DeepFeatIoT slightly outperforms it in accuracy. Notably, DeepHeteroIoT [Inan *et al.*, 2023] also exhibits competitive performance on the Urban dataset, ranking second in accuracy and third in F1 score, which indicates a close competition among these models.

In the Iowa dataset, DeepFeatIoT achieves an exceptional accuracy and F1 score of 99.33, outperforming the second-best models (InceptionTime [Ismail Fawaz *et al.*, 2020] and H-InceptionTime [Ismail-Fawaz *et al.*, 2022]) by nearly 2.6%. It is worth emphasizing that while ensembles of convolution-based architectures (*e.g.*, InceptionTime [Ismail Fawaz *et al.*, 2020] and H-InceptionTime [Ismail-Fawaz *et al.*, 2022]) can yield impressive results on well-balanced datasets, their performance often deteriorates on datasets with imbalanced class distributions (*e.g.*, Swiss).

In summary, our experimental results indicate that DeepFeatIoT is the most effective model when evaluated across four diverse IoT sensor datasets, as evidenced by its superior average accuracy and F1 scores. To assess efficiency, the average runtime (training and testing) of DeepFeatIoT across all datasets is approximately 10.8 minutes, which is commendable given its intricate architecture combining learned and non-learned feature extractors, yet remaining feasible for real-world use. Benchmark models such as DeepHeteroIoT [Inan *et al.*, 2023] and MultiROCKET [Tan *et al.*, 2022] are the next best alternatives in terms of consistency, while the previous state-of-the-art ensemble ML benchmark, MACE [Montori *et al.*, 2023], demonstrates competitive performance with an average accuracy of 91.80% and a F1

	Swiss	Urban	Iowa	SBAS
RF	78.1±0.4	87.3±0.4	83.3±1.2	97.4±0.0
PF	78.6±1.1	90.8±0.5	88.6±0.4	99.9±0.1
RF & PF	78.8±0.8	86.2±0.5	88.2±1.5	98.7±0.0
DC	77.4±1.2	90.3±1.3	83.2±3.3	98.7±0.0
Ours	89.7±0.9	96.5±0.3	98.7±0.4	98.9±0.5

Table 3: Ablation analysis of our proposed model based on mean and standard deviation of accuracy scores

²<http://newcastle.urbanobservatory.ac.uk/>

³<https://mesonet.agron.iastate.edu/ASOS/>

	Swiss	Urban	Iowa	SBAS
RF	71.9±0.4	81.2±0.9	82.7±1.1	97.4±0.0
PF	72.6±2.4	81.1±2.0	87.8±0.9	99.9±0.1
RF & PF	73.4±1.5	75.5±2.9	88.2±1.5	98.7±0.0
DC	71.1±1.8	82.3±3.2	82.8±3.9	98.7±0.0
Ours	88.7±1.2	93.9±0.4	98.7±0.4	98.9±0.5

Table 4: Ablation analysis of our proposed model based on mean and standard deviation of macro-average F1 scores

score of 90.32%. In contrast, transformer-based models like PatchTST [Nie *et al.*, 2022] perform poorly in the IoT time series classification domain, with average scores of 81.44% accuracy and 77.22% F1. Nevertheless, larger transformer architectures (e.g., GPT4TS [Zhou *et al.*, 2024]) that benefit from pre-training, exhibit promising potential by achieving perfect scores on the SBAS dataset and competitive overall averages. Moreover, DeepFeatIoT, PatchTSMixer [Ekambaram *et al.*, 2023], and DeepHeteroIoT [Inan *et al.*, 2023] all achieve perfect scores (100%) in both accuracy and F1 on the SBAS dataset.

4.2 Ablation Study

In this section, we present a detailed ablation study for major components of our proposed DeepFeatIoT model, to validate the effectiveness of our proposed deep learning model across datasets. Tables 3 and 4 respectively delineate the mean \pm standard deviation of accuracy and F1 scores over 10 runs. In both tables, the ablation of the proposed model is broken down into four parts: using only randomized features (RF), using only pre-trained LLM features (PF), using both randomized features and pre-trained LLM features (RF & PF), and finally, the results of direct concatenation (DC), which includes all learned and non-learned features but excludes our designed DFT module for improved feature transformation to scale the concatenation of dense features before the MLP classifier head.

It is evident from the mean accuracy and F1 scores presented in Tables 3 and 4 that our final proposed model which is an optimized unification of learned and non-learned features outperforms all other configurations. However, excluding the DFT module causes a significant drop in generalizability, as indicated by lower accuracy and F1 scores on three out of the four datasets. While configurations using only RF or only PF yield competitive performance, they do not consistently achieve superior results across all datasets. Furthermore, the combination of RF & PF, when employed without the inclusion of learned local and global features and the DFT module, fails to deliver competitive performance.

To statistically quantify the practical significance of our model’s improvements, we computed Cohen’s d statistics⁴ [Cohen, 2013] for mean accuracy and F1 score comparisons. The results indicate strong effect sizes across all ablation

⁴Following Cohen’s guidelines [Cohen, 2013], effect sizes are interpreted as: Cohen’s d statistic ≥ 0.8 (large, substantial improvement), indicating strong practical relevance of the observed performance gains.

variants, with values exceeding 1.0 in all cases, reinforcing the substantial performance gains achieved by our proposed model. Specifically, DeepFeatIoT demonstrates a significantly higher generalization capability, with Cohen’s d values reaching 1.44 (acc) and 1.43 (F1) against the RF-only variant, and consistently exceeding 1.0 across all comparisons. Overall, the computed Cohen’s d effect sizes for mean accuracy indicate substantial performance differences: Ours vs RF (1.443), Ours vs PF (0.939), Ours vs RF & PF (1.216), and Ours vs DC (1.191). Likewise, the effect sizes for F1 scores further reinforce this trend, with values of Ours vs RF (1.433), Ours vs PF (1.098), Ours vs RF & PF (1.231), and Ours vs DC (1.299), highlighting the significant impact of our model’s unified feature representation.

Furthermore, these results validate that the unified representation of features of our model - particularly the unification of learned local and global features with randomized and pre-trained LLM features - plays a critical role in improving classification performance in diverse IoT sensor datasets. Furthermore, the large effect sizes (Cohen’s $d \geq 0.8$) confirm that the observed performance gains are not only statistically significant but also practically meaningful for real-world IoT sensor classification. Additionally, our experimental observations and analysis also suggest that state-of-the-art pre-trained LLM architectures originally designed for text generation, such as GPT2 [Radford *et al.*, 2019], can be effectively incorporated for feature extraction from raw IoT time series even without domain-specific adapter blocks like previous studies [Zhou *et al.*, 2024] or re-programming [Nie *et al.*, 2022] or transformation (e.g., frequency domain conversion) of raw time series, paving the way for future research in heterogeneous IoT sensor data classification for critical application domains.

5 Conclusion

While previous state-of-the-art models often struggle with generalization or overfitting, particularly due to challenges like limited labelled data and class imbalance, our proposed deep learning model effectively addresses these issues without requiring additional data preprocessing or augmentation, even for smaller IoT sensor datasets (e.g., Swiss). The automated classification of IoT sensor data eliminates the need for manual annotation from heterogeneous sources, paving the way for smarter critical applications across various industries. If deployed in real-world industrial settings, the proposed model would significantly reduce manual labor while saving both financial costs and valuable time across multiple critical sectors. Ultimately, this approach promotes the reusability of vast amounts of meaningful IoT sensor data, enabling advancements in critical technologies.

Acknowledgments

This research was funded by the Australian Research Council (ARC) through an ARC Discovery Project grant (DP220101420).

References

- [Bagnall *et al.*, 2017] Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data mining and knowledge discovery*, 31:606–660, 2017.
- [Borges *et al.*, 2022] João B Borges, Heitor S Ramos, and Antonio AF Loureiro. A classification strategy for internet of things data based on the class separability analysis of time series dynamics. *ACM Transactions on Internet of Things*, 3(3):1–30, 2022.
- [Calbimonte *et al.*, 2012] Jean-Paul Calbimonte, Oscar Corcho, Zhixian Yan, H Jeung, and Karl Aberer. Deriving semantic sensor metadata from raw measurements. In *Proceedings of the 5th International Workshop on Semantic Sensor Networks*, 2012.
- [Cohen, 2013] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. routledge, 2013.
- [Čulić Gambiroža *et al.*, 2023] Jelena Čulić Gambiroža, Toni Mastelić, Ivana Nižetić Kosović, and Mario Čagalj. Lost in data: recognizing type of time series sensor data using signal pattern classification. *International Journal of Data Science and Analytics*, pages 1–12, 2023.
- [Dempster *et al.*, 2020] Angus Dempster, François Petitjean, and Geoffrey I Webb. Rocket: exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495, 2020.
- [Ekambaram *et al.*, 2023] Vijay Ekambaram, Arindam Jati, Nam Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 459–469, 2023.
- [Georgakopoulos *et al.*, 2020] Dimitrios Georgakopoulos, Prem Prakash Jayaraman, and Anas Dawod. Senshamart: A trusted lot marketplace for sensor sharing. In *2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC)*, pages 8–17, 2020.
- [Gruver *et al.*, 2024] Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew G Wilson. Large language models are zero-shot time series forecasters. *Advances in Neural Information Processing Systems*, 36, 2024.
- [Hong *et al.*, 2017] Dezhi Hong, Quanquan Gu, and Kamin Whitehouse. High-dimensional time series clustering via cross-predictability. In *Artificial Intelligence and Statistics*, pages 642–651. PMLR, 2017.
- [Inan *et al.*, 2023] Muhammad Sakib Khan Inan, Kewen Liao, Haifeng Shen, Prem Prakash Jayaraman, Dimitrios Georgakopoulos, and Ming Jian Tang. Deepheteroiot: Deep local and global learning over heterogeneous iot sensor data. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 119–135. Springer, 2023.
- [Ismail Fawaz *et al.*, 2020] Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F Schmidt, Jonathan Weber, Geoffrey I Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962, 2020.
- [Ismail-Fawaz *et al.*, 2022] Ali Ismail-Fawaz, Maxime Devanne, Jonathan Weber, and Germain Forestier. Deep learning for time series classification using new hand-crafted convolution filters. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 972–981. IEEE, 2022.
- [Jiménez and Raj, 2019] Abelino Jiménez and Bhiksha Raj. Time signal classification using random convolutional features. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3592–3596. IEEE, 2019.
- [Khan *et al.*, 2022] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41, 2022.
- [Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [Liu *et al.*, 2021] Mei Liu, Liangming Chen, Xiaohao Du, Long Jin, and Mingsheng Shang. Activated gradients for deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 34(4):2156–2168, 2021.
- [Manoharan *et al.*, 2021] Kalaiselvi Geetha Manoharan, Jawaharlal Arun Nehru, and Sivaraman Balasubramanian. Artificial intelligence and iot. *Springer*, 10:978–981, 2021.
- [Min *et al.*, 2023] Bonan Min, Hayley Ross, Elinor Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, 56(2):1–40, 2023.
- [Montori *et al.*, 2023] Federico Montori, Kewen Liao, Matteo De Giosa, Prem Prakash Jayaraman, Luciano Bononi, Timos Sellis, and Dimitrios Georgakopoulos. A metadata-assisted cascading ensemble classification framework for automatic annotation of open iot data. *IEEE Internet of Things Journal*, pages 1–1, 2023.
- [Nie *et al.*, 2022] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2022.
- [Postol *et al.*, 2019] Michael Postol, Candace Diaz, Robert Simon, and Drew Wicke. Time-series data analysis for classification of noisy and incomplete internet-of-things datasets. In *2019 18th IEEE International Conference*

On Machine Learning And Applications (ICMLA), pages 1543–1550, 2019.

[Radford *et al.*, 2019] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[Rajak and Selot, 2021] Ranjit Rajak and Kritika Selot. Internet of things (iot): State-of-the-art technologies, challenges and applications. *Machine Intelligence and Smart Systems: Proceedings of MISS 2020*, pages 117–130, 2021.

[Stoyanova *et al.*, 2020] Maria Stoyanova, Yannis Nikoloudakis, Spyridon Panagiotakis, Evangelos Pallis, and Evangelos K Markakis. A survey on the internet of things (iot) forensics: challenges, approaches, and open issues. *IEEE Communications Surveys & Tutorials*, 22(2):1191–1221, 2020.

[Tan *et al.*, 2022] Chang Wei Tan, Angus Dempster, Christoph Bergmeir, and Geoffrey I Webb. Multirocket: multiple pooling operators and transformations for fast and effective time series classification. *Data Mining and Knowledge Discovery*, 36(5):1623–1646, 2022.

[Wajgi *et al.*, 2023] Rakhi Wajgi, Jitendra V Tembhurne, Dipak Wajgi, and Tapan Jain. Internet of everything: Applications. In *Modern Approaches in IoT and Machine Learning for Cyber Security: Latest Trends in AI*, pages 131–157. Springer, 2023.

[Yang *et al.*, 2024] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Shaochen Zhong, Bing Yin, and Xia Hu. Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data*, 18(6):1–32, 2024.

[Zhou *et al.*, 2024] Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36, 2024.