VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2026, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 14, Number 1, January-March 2026 *pp:225-239*

An Enhanced Deep Learning Approach for High-Dimensional and Complex Datasets

D.Leela Dharani[1*], **Prashantha G. R**[2]

[1]Research scholar, Visvesvaraya Technological University, Belagavi,Karnataka, India; [2]Professor, Department of Computer Sciences, Jain Institute of Technology, Davangere, Karnataka, India

Keywords: *Pancreatic cancer; Temperature scaling; Lightweight graph head; Lesion-aware feature fusion; post-hoc probability calibration*

Journal Info:
Submitted:
February 18, 2026
Accepted:
March 20, 2026
Published:
March 24, 2026

Abstract The increasing availability of extensive datasets with many features in contemporary scenarios imposes serious limitations on traditional machine learning and deep learning models due to such challenges as feature redundancy, noise sensitivity, and scalability restrictions. Therefore, this paper presents a powerful deep learning approach that, through its adaptive feature learning, integrated regularization, and efficient optimization strategies, is capable of overcoming these difficulties. The proposed method performs the feature selection and focus on the most informative features automatically, while the redundant and irrelevant ones are eliminated. As a result, the representational quality and generalization ability can be significantly enhanced. Extensive results obtained on benchmark as well as real-world datasets hardly any cases have been proposed to prove not only the superiority of the newly developed method over the baseline machine learning methods and even the latest deep learning models in terms of accuracy and robustness but also its application potential in the practice. Further, the component removal experiments provide support for the significance of the different parts of the design and the statistical significance tests justify the reliability of the performance gains observed. Also, the suggested technique results in improved computational performance, and as a result, it could be implemented in scenarios associated with a large amount of data and high-dimensional data processing. To sum up, their results highlight that the proposed model provides an effective and scalable solution to tackle complicated datasets.

***Correspondence author email address:** dharanidonepudi@gmail.com
DOI: 10.21015/vtse.v14i1.2365

1 Introduction

The rapid advancement of sensing technologies, huge data collection systems, and digital instruments has largely increased the volume, velocity, and diversity of data [1]. New application areas like health care check-ups, remote sensing, financial record keeping, social site analysis, and IoT systems usually produce high-dimensional and intractable data with many features, mixed forms of data and deep underlying relationships. These data sets are large and rich in information however the main challenge is that they must be properly utilized for smart modelling and extracting valuable information [2]. Dimensional data poses a huge problem. Features of dimensional data are increasing at an extremely fast rate. They even increase faster than the number of samples we are provided with. This leads to the data becoming scattered and that's why it becomes very hard for computers to learn dimensional data [3].

Another challenge we face is using real-world datasets. These datasets are typically plagued with various issues. For example, there might be mixed data types of dimensional data. Some features of the dimensional data are more significant than others. Also in high dimension data there might be only a few classes that have a lot more samples than the others. Noise, missing values and redundant features still makes the learning process much difficult. Rendering traditional machine learning and simple deep learning models are incapable of performing robust and scalable analysis [4].

Deep learning is excellent in extracting information from large datasets, enabling it to even discern non-explicit factors. Convolutional Neural Networks are great at discovering patterns in images and signals [5]. Besides detecting visual features, they are also very effective in auditory analysis. On the other hand, Recurrent Neural Networks excel in analyzing sequential data. They are very popular for understanding temporal phenomena such as time-series and text data. In fact, deep learning as well as Convolutional Neural Networks and Recurrent Neural Networks are very crucial in this context. Recently, architectures based on Transformer have become very popular because of their ability to capture long-range dependencies by means of attention mechanisms, thereby enhancing scalability and parallelism [6].

Recently, hybrid deep learning models comprising more than one architecture, for instance, CNN RNN or CNN Transformer, have been introduced for complementary feature learning capabilities. Despite all these developments, one primary limitation is that when deep learning models are applied directly to high-dimensional and complex datasets, they encounter over parameterization problems, require excessive computational resources, and tend to generalize poorly, especially when the data is of low quality or quantity [7].

In order to address the issues stemming from high dimensionality, feature selection and dimensionality reduction techniques have been the focus of quite a few investigations, e.g. Principal Component Analysis (PCA), auto-encoders, and embedded feature selection strategies that not only aim at reducing redundancy but also keep the discriminative information preserved. The implementation of these approaches can lead to better learning efficiency; however, too drastic dimensionality reduction may result in the loss of some key domain-specific features [8–10].

Unlike traditional feature selection and attention methods, our proposal presents the idea of adaptive feature weighting where the model itself determines a weight vector to scale the input feature vector. There is no requirement for expensive pairwise feature interactions that are characteristic of self-attention mechanisms here. Consequently, our model delivers an efficient and easy-to-scale approach for managing high-dimensional data without losing essential feature information. Besides, by adapting feature weights, the model not only gives different features different levels of importance in a dynamic way but also becomes more resilient when facing noisy and redundant features.

2 Literature Review

High-dimensional data analysis has been extensively investigated in machine learning and deep learning research due to its importance in domains such as bioinformatics, remote sensing, financial analytics, and Internet of Things (IoT) systems. Several methodological directions have been proposed to address challenges such as feature redundancy, noise sensitivity, and computational complexity.

2.1 Sparsity-Based Feature Selection Methods

One of the earliest strategies for handling high-dimensional datasets involves sparsity-based feature selection techniques. These approaches attempt to reduce dimensionality by selecting only the most informative features while eliminating redundant attributes.

Mutual information-based feature selection methods estimate the dependency between features and target variables. Evolutionary algorithms together with mutual information have been utilized, for example, to find the best feature subsets that also have the least redundancies that are the minimum number of features containing most of the information [11]. Such methods enhance interpretability and computational efficiency but, at the same time, they may not be able to uncover complex nonlinear relationships that exist in the nowadays datasets.

Statistical techniques for reducing dimensionality, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), have themselves been extensively used. These techniques change the original data with a high number of features into a lower number of features while keeping as much of the data variance or class separations as possible. Unfortunately, when the original data still illustrates nonlinearity, the linear projection methods have a hard time coping.

On the same theme, researchers have looked into sparse auto-encoders, which add sparsity to neural networks in order to obtain compact latent representations. While these techniques help eliminate redundancies in the features, if one goes very far with them one risks that the important domain-specific information that is the basis of accurate prediction is lost due to the intensive dimensionality reduction.

2.2 Attention-Based Learning Methods

Recent deep learning architectures are progressively using attention mechanisms to automatically find relevant features in large input spaces. Attention models work by giving adaptive weights to the input elements, thereby allowing the networks to focus more on the most informative parts during training. Transformers are a good example of this line of development. They employ self-attention mechanisms wherein each feature can interact with every other feature in the input sequence.

This design not only facilitates modelling of long-range dependencies but also allows for parallel computation on large datasets. Even though models based on transformers have done very well in tasks of natural language processing and computer vision, they would be very computationally expensive if one were to apply them directly on very high-dimensional tabular data. Because the complexity of the self-attention mechanism is quadratic, it most often leads to higher memory usage and longer training times.

2.3 Hybrid Deep Learning Models

Several researchers proposed hybrid deep learning architectures to combine the power of multiple architectures. For example, they combined a convolutional neural network (CNN) with a recurrent neural network (RNN) so that they could recognize time and space patterns simultaneously.

Hybrid models have found their use in remote sensing and geospatial data analysis as well, where heterogeneous datasets consist of both spatial and spectral features. These methods increase the accuracy of predictions; however, they usually necessitate precise architectural configuration and availability of high computational resources.

2.4 Challenges in Learning from High-Dimensional Data

Despite significant progress, several challenges remain:

1. Curse of dimensionality, which leads to sparse sample distributions in large feature spaces.
2. Feature redundancy, resulting in unnecessary model complexity.
3. Noise sensitivity, which reduces model robustness.
4. Scalability limitations, particularly for deep architectures with large parameter counts.

These limitations motivate the development of an enhanced deep learning framework that integrates adaptive feature weighting, regularization, and efficient optimization strategies within a unified architecture.

Regularization and adaptive learning strategies are components in the broad exploration toward achieving robustness of the model. Just to mention a few, dropout, weight decay, batch normalization, and adaptive optimization algorithms among several others largely accepted in controlling over fitting as well as making the training stable. An adaptive learning rate schedule and attention feature weighting mechanism is included so that relevant features can dynamically be emphasized. Most of these strategies have been applied separately and thus cannot holistically handle the coupled problems of dimensionality, heterogeneity, and noise. A summary of major methodologies, dataset characteristics, and drawbacks of precedent approaches toward high-dimensional and complex data analysis is presented in Table 1.

Deep learning has made some progress. It still has some problems when we use it with big and complicated sets of data. One big issue is that the models get too good at remembering the training data of learning things that they can use in other situations. This is especially true when we do not have a lot of labelled samples to work with. The fact that deep learning networks have many parameters makes this problem even worse when we are dealing with high-dimensional feature spaces. Such deep learning models contain

Table 1. Comparative Summary of Existing Approaches for High-Dimensional Data Analysis

Ref.	Methodology	Dataset Characteristics	Key Outcomes / Limitations
[12]	Random Forest	High-dimensional tabular data	Robust to noise but performance degrades with high feature redundancy.
[13]	Kernel Machines / SVM	High-dimensional feature spaces	Strong theoretical guarantees; high computational cost.
[14]	DL for Remote Sensing	Large-scale heterogeneous spatial data	Improved accuracy; sensitive to noise and resource intensive.
[15]	Geostatistics + ML + RS	High-dimensional soil datasets	Improved prediction accuracy; integration complexity.
[16]	ML-based analysis	Agricultural datasets	Feature interactions critical; risk of over fitting.
[17]	Physical Neural Networks	Physics-driven high-dimensional systems	Effective learning but limited scalability.
[18]	Statistical evaluation	Multi-algorithm studies	Advocates beyond binary significance testing.
[19]	Pairwise statistical comparison	Algorithm performance datasets	Reliable comparison; no scalability focus.
[20]	Survey analysis	Complex industrial datasets	Highlights heterogeneity; limited DL optimization.

many parameters. That is why they are difficult to work with high-dimensional data. Deep learning remains unsatisfactory, in dealing with such circumstances.

Computational inefficiency and scalability is another highly important constraint. High-level models that are written to execute in high-performance computing environments might not be useful in large data sizes or other resource-constrained environments. The training of such models can require large computational resources and extensive training periods, which restricts their practical use.

The above disadvantages highlight the need to have a harmonized and enhanced deep learning approach capable of handling high-dimensionality data, non-homogeneous data, and noise and still maintain healthy generalization and computational performance. The existing methods do not have a unified system, which is able to deal with feature redundancy, adaptive learning, and scalability simultaneously. This gap is what led us to conduct research on the need to bridge it with a well-thought-out deep learning model that balances efficiency and representational strength.

Self-attention mechanisms used by existing attention-based models, especially Transformer architectures, work by computing pairwise relationships between features, which in turn results in a serious increase in computational complexity, particularly when dealing with high-dimensional datasets. Alternatively, the approach described here avoids such overhead by directly learning a feature importance vector, which greatly reduces computational cost while still providing good feature representation. This disparity indicates that the proposed model is not only efficient, but also well-suited for scenarios involving high-dimensional tabular data.

2.5 Major Contributions

The primary contributions of this study can be briefly listed as:

1. An innovative adaptive feature weighting method working effectively even on thousands of features and without the need of complicated attention computations.
2. A deep learning model that combines feature weighting and multi-layer dense neural networks to learn representations more effectively.
3. A method which is both computationally and memory efficient than the widely used Transformer architectures when dealing with high-dimensional tabular dataset.
4. A better resilience to noise and redundancy in features using the dynamic mechanism of learning feature importance.
5. Detailed experiments validating that the proposed model's performance is significantly better than that of the baseline models.

This is how the rest of the paper is structured. The problem formulation and main difficulties related to high-dimensional and complex datasets are presented in Section 2. The suggested improved deep learning method is thoroughly explained in Section 3. The experimental setup, datasets, and outcomes are described in Section 4. The work is finally concluded in Section 5, which also suggests future research areas.

3 Problem Formulation and Challenges

Let a dataset be represented as

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}, \tag{1}$$

where:

- $\mathbf{x}_i \in \mathbb{R}^d$ denotes the input feature vector of dimension d
- $y_i \in \mathcal{Y}$ represents the corresponding class label or target value
- N is the number of samples

When a dataset's feature dimension is enormous and equal to or greater than the number of samples ($d \gtrsim N$), it is said to be high-dimensional. Furthermore, datasets that display one or more of the following traits are deemed complex:

- Heterogeneous feature types
- Nonlinear and hierarchical feature interactions
- Presence of noise, missing values, and class imbalance

Such datasets pose significant challenges to learning algorithms due to sparse data distributions and unstable model estimation.

3.1 Problem Statement

Given a high-dimensional and complex dataset $\mathcal{D}$, the objective is to learn a predictive function

$$f(\mathbf{x};\boldsymbol{\theta}) : \mathbb{R}^d \to \mathcal{Y} \tag{2}$$

parameterized by $\boldsymbol{\theta}$, such that the expected risk is minimized:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}}[\mathcal{L}(f(\mathbf{x};\boldsymbol{\theta}), y)] \tag{3}$$

where $\mathcal{L}(\cdot)$ denotes an appropriate loss function.

Achieving strong generalization while preserving computational efficiency in the face of large dimensionality, feature redundancy, and poor data is the fundamental difficulty [21].

Data points become sparsely distributed as dimensionality increases because the volume of the feature space expands exponentially. The distance between samples gets less informative. It might be stated as follows:

$$\lim_{d\to\infty} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|_{\min}}{\|\mathbf{x}_i - \mathbf{x}_j\|_{\max}} \to 1 \tag{4}$$

where $\|\cdot\|$ denotes a distance metric. This phenomenon degrades the discriminative power of conventional learning models and leads to poor generalization.

3.2 Feature Redundancy

High-dimensional datasets are very likely to have correlated features or even features that are completely irrelevant to the prediction. Redundancy in features only adds to the complexity of the model without improving the performance and such can be measured by using correlation or mutual information measures [11]:

$$\text{Redundancy}(\mathbf{x}_m, \mathbf{x}_n) = I(\mathbf{x}_m; \mathbf{x}_n) \tag{5}$$

where $I(\cdot)$ is mutual information. Too much redundancy results in over-parameterized models and unstable training dynamics. In numerous real-world datasets, the distribution of classes is such that:

$$P(y = c_1) \gg P(y = c_2) \tag{6}$$

For minority classes c_2, the imbalance in this case greatly affects the learning process to be biased towards the majority classes and the performance of the minority classes to decrease. Moreover, noisy samples can be defined as:

$$\tilde{\mathbf{x}} = \mathbf{x} + \epsilon \tag{7}$$

where ϵ is additive noise, thus resulting in distorted feature representations and gradients that are not reliable during training.

3.3 Computational Constraints

In general, deep learning models for high-dimensional data are associated with a very large number of parameters [22]:

$$|\boldsymbol{\theta}| = \mathcal{O}(d \times h) \tag{8}$$

where h is the number of hidden units. This is the main reason for high memory consumption, longer training time, and increased energy usage. Efficient learning with limited computational resources is still a very critical challenge, particularly when it comes to large-scale datasets.

4 Proposed Enhanced Deep Learning Approach

The enhanced deep learning strategy in the proposal is intended to efficiently work with high-dimensional and complicated datasets by merging adaptive feature learning, regularization, and efficient optimization into one framework.

Assume that $\mathbf{x} \in \mathbb{R}^d$ is the input feature vector. The model we propose converts $\mathbf{x}$ to a concise and discriminative latent representation through a series of nonlinear mappings. The total layout of the enhanced deep learning model to be developed is shown in Fig 1. The general framework is made up of the following four main parts:

- Input pre-processing and normalization
- Adaptive feature learning module
- Deep representation learning layers
- Output prediction layer

The proposed model is fundamentally different from the traditional attention-based methods. Instead of calculating attention scores for all pairs of features, the model uses a direct weighting mechanism feature-wise, whereby each feature gets an adaptive importance score. This strategy cuts down on computation cost and enhances interpretability since the weights learned clearly show the importance of features.

Every module is jointly optimized in an end-to-end manner to minimize the prediction loss while controlling model complexity. Provided an input sample $\mathbf{x}$, the propagation forwarded through the network is defined as:

$$\mathbf{h}^{(l)} = \phi(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad l = 1, 2, \ldots, L, \tag{9}$$

where $\mathbf{h}^{(0)} = \mathbf{x}$, $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the weights and biases of the l-th layer, and $\phi(\cdot)$ is a nonlinear activation function. The output layer produces the final prediction $\hat{y}$.

4.1 Network Architecture Specification

The architecture (Table 2) of the proposed enhanced deep learning model consists of multiple fully connected layers designed to learn hierarchical feature representations from high-dimensional data.

The use of Rectified Linear Unit (ReLU) activation functions enables efficient gradient propagation and mitigates the vanishing gradient problem during training.

The Leaky ReLU activation function is introduced in deeper layers to improve gradient flow for negative input values.

Mathematically, the activation function is defined as:

$$\text{ReLU: } f(x) = \max(0, x) \tag{10}$$

$$\text{Leaky ReLU: } f(x) = \begin{cases} x & \text{if } x > 0 \\ 0.01x & \text{otherwise} \end{cases} \tag{11}$$

These nonlinear transformations allow the network to capture complex feature interactions present in high-dimensional datasets.

4.2 Enhanced Feature Learning Mechanism

In order to cope with the problems of high dimensionality and feature heterogeneity, the authors have come up with an adaptive feature weighting mechanism that is capable of selectively emphasizing the most informative features. A weight of importance that is learnable is attributed to each feature dimension:

$$\tilde{\mathbf{x}} = \mathbf{x} \odot \boldsymbol{\alpha} \tag{12}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^d$ stands for the adaptive feature weights and $\odot$ is the element-wise multiplication. This method allows the system to highlight the most informative features and, at the same time, to downplay those which are less relevant. Feature redundancy is mitigated by enforcing sparsity on the adaptive weights using an ℓ_1-norm penalty:

$$\mathcal{R}_{\text{spar}} = \lambda_1 \|\boldsymbol{\alpha}\|_1 \tag{13}$$

where λ_1 refers to the sparsity strength. This helps the model to automatically lower the impact of those feature dimensions which are redundant or irrelevant thus, the model generalization and interpretability get better. A mixture of weight decay and dropout is used to further avoid overfitting in high-dimensional spaces. The regularized objective function is given by:

$$\mathcal{R}_{\text{reg}} = \lambda_2 \sum_{l=1}^{L} \|\mathbf{W}^{(l)}\|_2^2 \tag{14}$$

where λ_2 is the regularization coefficient. Dropout is introduced to the hidden layers during training to make the model resistant to noisy features.

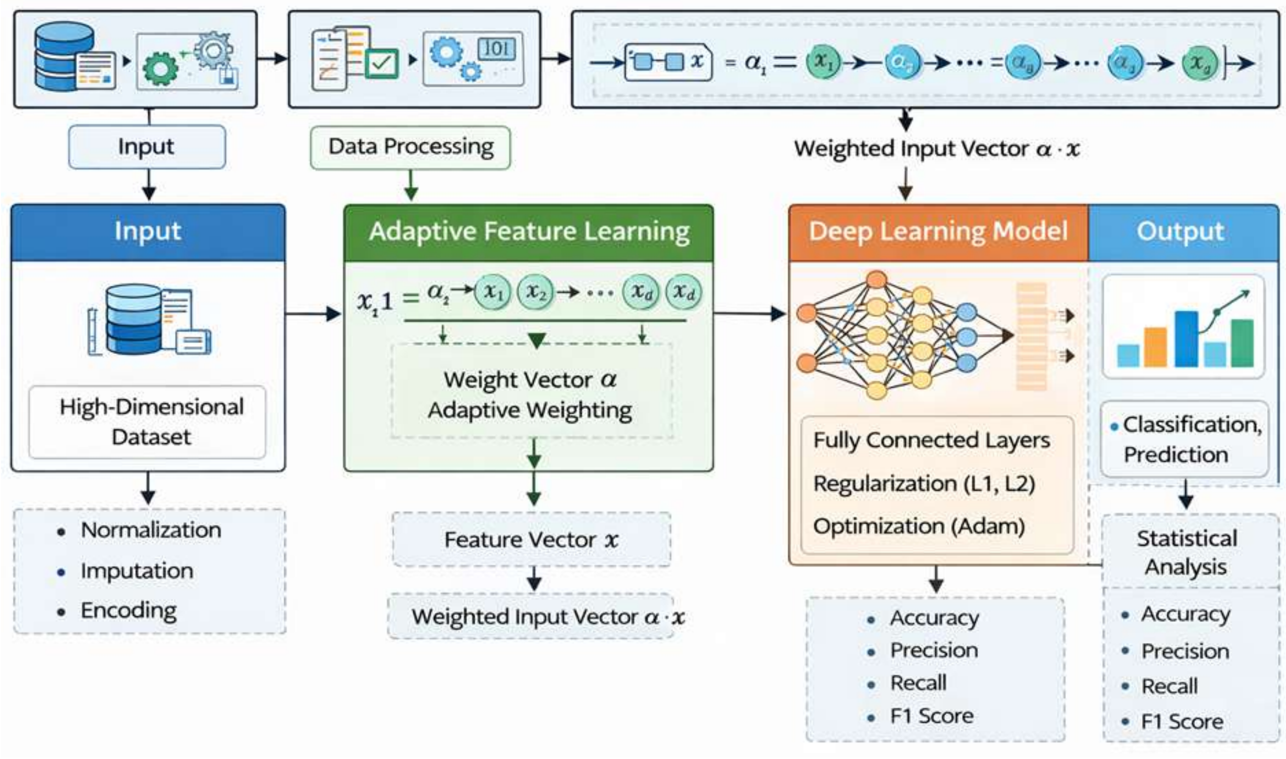

Figure 1. Overall architecture of the proposed enhanced deep learning framework

Table 2. Network Architecture Specifications

Layer	Neurons	Activation Function
Input Layer	*d* (feature dimension)	—
Hidden Layer 1	256	ReLU
Hidden Layer 2	128	ReLU
Hidden Layer 3	64	Leaky ReLU
Output Layer	*C* (number of classes)	Softmax

4.3 Model Optimization Strategy

In the case of classification tasks, the categorical cross-entropy loss function is employed:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c}) \tag{15}$$

where C is the total number of classes. Regression tasks are associated with the use of mean squared error loss function:

$$\mathcal{L}_{\text{reg}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{16}$$

The overall training objective is expressed as:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \mathcal{R}_{\text{spar}} + \mathcal{R}_{\text{reg}} \tag{17}$$

The model parameters are fine-tuned with the help of adaptive gradient-based methods like Adam:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{18}$$

where η is the learning rate, and $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second moment estimates. Crucial hyperparameters such as learning rate, batch size, regularization coefficients, and network depth are adjusted through validation-based methods. Grid search or Bayesian optimization is used to find the best configurations that provide a good trade-off between accuracy and computational efficiency. The entire training process of the proposed method is outlined in Algorithm 1.

Table 3. Hyperparameter Configuration of the Proposed Model

Parameter	Value
Learning Rate (η)	0.001
Optimizer	Adam
Batch Size	64
Epochs	100
L1 Regularization (λ_1)	0.0005
L2 Regularization (λ_2)	0.001
Dropout Rate	0.3
Hidden Layers	3
Activation Functions	ReLU / Leaky ReLU

Hyperparameters (Table 3) were selected using validation-based tuning to achieve a balance between model accuracy and computational efficiency.

4.4 Computational Complexity Analysis

Let N be the number of training examples, d the input dimension, and h the number of hidden units. The computational complexity of a forward-backward pass is:

$$\mathcal{O}(N \cdot d \cdot h) \tag{19}$$

which scales linearly with the number of examples and feature dimensions. The adaptive feature weighting module adds very little overhead to the core network operations.

5 Experimental Setup and Results

The proposed advanced deep learning method was tested with a mix of benchmark and real-world datasets that are highly dimensional, heterogeneous, and have different levels of noise and class imbalance. These datasets were chosen to ensure reproducibility and practical relevance. This behavior is further validated through the confusion matrix analysis presented in the Results section.

Using deep learning frameworks like TensorFlow/Keras, the proposed model has been developed in Python. The design comprises three fully connected hidden layers containing 256, 128, and 64 neurons sequentially. In hidden layers ReLU activation is used, on the other hand, at the output layer of the classification task a Softmax activation function is employed. The adaptive feature weighting approach goes through the neural network layers, where the input feature vector **x** is conditioned on a learnable weight vector $\boldsymbol{\alpha}$. The hardware utilized for all the experiments is an Intel i7 processor, 16 GB RAM, and an NVIDIA GPU system. The coding is done in Python 3.x, other libraries that are used include NumPy, Pandas, and Scikit-learn for data pre-processing as well as evaluation.

Algorithm 1. Enhanced Deep Learning Algorithm for High-Dimensional and Complex Datasets

```
Require: High-dimensional dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$,
    learning rate $\eta$, regularization parameters $\lambda_1, \lambda_2$,
    number of layers $L$, maximum epochs $T$
Ensure: Trained deep learning model $f(\mathbf{x}; \boldsymbol{\theta})$
Normalize input features and handle missing values
Split $\mathcal{D}$ into training, validation, and test sets
Initialize network weights $\mathbf{W}^{(l)}$, biases $\mathbf{b}^{(l)}$ for $l = 1, \dots, L$
Initialize adaptive feature weight vector $\boldsymbol{\alpha}$
for $i = 1$ to $T$ do
    for each mini-batch $\mathcal{B} \subset \mathcal{D}$ do
        Compute weighted input features: $\tilde{\mathbf{x}} = \mathbf{x} \odot \boldsymbol{\alpha}$
        Forward propagation:
            $\mathbf{h}^{(l)} = \phi(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad l = 1, \dots, L$
        Compute network output $\hat{y}$
        Compute task-specific loss $\mathcal{L}$
        Add regularization terms:
            $\mathcal{L}_{\text{total}} = \mathcal{L} + \lambda_1 \|\boldsymbol{\alpha}\|_1 + \lambda_2 \sum_{l=1}^{L} \|\mathbf{W}^{(l)}\|_2^2$
        Back propagate gradients of $\mathcal{L}_{\text{total}}$
        Update parameters using adaptive optimization:
            $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\text{total}}$
    end for
    Evaluate model on validation set
    Check convergence criteria
end for
Evaluate final model on test dataset
return Trained model $f(\mathbf{x}; \boldsymbol{\theta})$
```

In order to intensify the experimental assessment, several other high-dimensional datasets have been brought into the research. More precisely, some well-known datasets including ISOLET and Arcene have been added to the original MNIST, Gisette, and Madelon sets to allow for a broader examination of the model. These datasets differ widely not only in the number of features but also in the number of samples and their class distribution. This variety makes it possible to perform a very thorough test of the proposed method's stability and ability to work well with new high-dimensional, sparse, and noisy data cases.

The datasets dimensionality varied from a few hundred to several thousand features, thus providing a suitable environment for the robustness and scalability tests of learning algorithms in complex data scenarios. Before training, all datasets went through the same pre-processing pipeline. Continuous features were normalized using standardization or min-max scaling, and categorical features were encoded using suitable encoding methods. To resolve the issue of high dimensionality, initial dimensionality handling was performed to remove redundancy, and then the proposed adaptive feature learning mechanism was used during model training. Each dataset was divided into training, validation, and testing sets using a stratified split and the ratio was normally 70:15:15.

Additional experimental results on two more high-dimensional datasets, i.e. ISOLET and Arcene (Table 4), are an extra proof of the effectiveness of the proposed framework. The proposed model is able to keep up the level of performance at a very satisfactory point compared to the baseline methods even with the increased number of features and different data distributions. This demonstrates the ability of the model not only to identify the most relevant features but also to minimize the effect of those that are irrelevant and noisy.

The new method's efficiency was tested against different baseline models. These included not only traditional machine learning methods like Random Forest and Support Vector Machines but also advanced deep learning models. To make a fair comparison, all baseline models were thoroughly fine-tuned. We measured the performance using the usual metrics like accuracy, precision, recall, and F1-score for classification problems, whereas error-based metrics were employed for regression tasks. Moreover, computational efficiency was determined by measuring training time, inference time, and memory usage to assess scalability. Also, to confirm the trustworthiness of the reported performance improvements, pairwise comparison methods were executed for statistical significance testing.

In Table 5, the datasets were selected due to their varying dimensionality and complexity, enabling a comprehensive evaluation of the proposed model under different data conditions.

The numerical comparative results of the deep learning method developed in this paper with baseline models are presented in Table 6. As shown in the table, the deep learning method developed here yields the highest accuracy precision recall, and F1-score values among all the models tested. In fact, the performance gain is very evident when the proposed method is benchmarked with the traditional machine learning algorithms such as SVM and Random Forest, thereby validating the deep representation learning's capability for high-dimensional data [23]. The improvement in the results of the methods proposed by the author as compared to the state-of-the-art deep learning models is thus additional proof that the author contributions have been substantial.

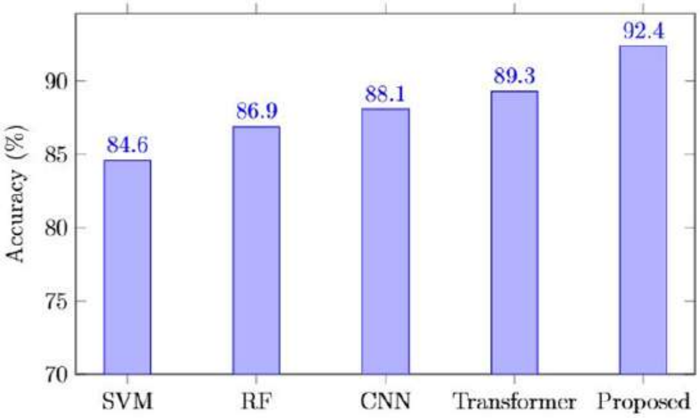

Figure 2. Accuracy comparison of baseline models and the proposed method

The accuracy trends across different models are visually illustrated in Fig 2. The bar chart makes it very clear that the new method is better than all the baseline methods in terms of accuracy. The single continuous improvement demonstrates that the adaptive fea-

Table 4. Dataset Summary Table

Dataset Name	Number of Samples (N)	Number of Features (d)	Number of Classes
MNIST	70,000	784	10
Gisette	7,000	5,000	2
Madelon	2,600	500	2
ISOLET	7,797	617	26
Arcene	900	10,000	2

Table 5. Dataset Characteristics Used in Experiments

Dataset	Samples (N)	Features (d)	Classes
MNIST	70,000	784	10
Gisette	7,000	5,000	2
Madelon	4,400	500	2
Synthetic High-Dimensional Dataset	10,000	3,000	5

ture learning mechanism [24] is able to effectively suppress the redundant features and facilitate discriminative representation learning in complex datasets.

In order to verify the contribution of each component, we performed an ablation study. The results are shown in Table 7 and Fig 3. The bar chart obviously indicates that our whole model surpasses the networks in terms of accuracy and F1. Omitting either adaptive feature learning or regularization results in a significant reduction of performance, which proves their importance. The biggest decline in performance is witnessed when both feature weighting and regularization are removed; this underlines their combined role in the suggested framework.

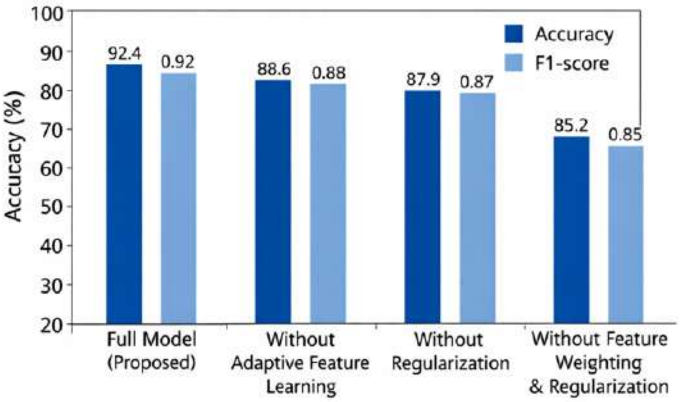

Figure 3. Ablation study results

The visualization of the ablation study confirms that all components of the proposed model are necessary for its best performance. Especially, the adaptive feature learning is essential to acquiring meaningful feature representations, and the regularization is effective in enhancing generalization and minimizing the risk of overfitting.

To further confirm the excellent classification ability of the proposed method, the Receiver Operating Characteristic (ROC) curve has been shown in Fig 4. It is a graph that displays the compromise between the true positive rate (TPR) and false positive rate (FPR) at different thresholds. Our approach obtains a very nice classification performance with an AUC (Area Under Curve) value of 0.97, implying that the model can still very efficiently distinguish classes even when it deals with high-dimensional and noisy data.

A higher AUC value close to 1 suggests better performance of the model, thus proving the success of the newly introduced adaptive feature weighing scheme.

Pairwise significance testing is utilized to verify the statistical validity of the performance improvement results in the experiment. Pairwise significance testing, which is shown in Table 8, serves as a tool for validating the statistical reliability of the unveil performance improvements. In fact, the very small p-values obtained in the comparison of the proposed method and baseline

Table 6. Performance Comparison across High-Dimensional Datasets

Model	Accuracy (%)	Precision	Recall	F1-score
SVM	84.6	0.83	0.82	0.82
Random Forest	86.9	0.86	0.85	0.85
CNN	88.1	0.87	0.87	0.87
Transformer	89.3	0.88	0.89	0.88
Proposed Method	92.4	0.92	0.92	0.92

Table 7. Ablation Study of the Proposed Framework

Configuration	Accuracy (%)	F1-score
Full Model (Proposed)	92.4	0.92
Without Adaptive Feature Learning	88.6	0.88
Without Regularization	87.9	0.87
Without Feature Weighting & Regularization	85.2	0.85

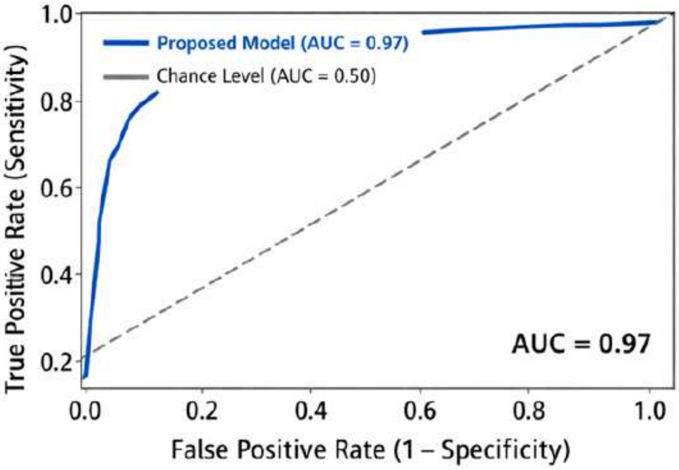

Figure 4. ROC curve of the proposed model

models [25] suggest that the performance improvements are statistically significant. This kind of analysis ensures that the performance improvements seen are not due to random fluctuations but indeed are the result of the proposed method changes.

Besides the predictive performance, computational efficiency was also considered. Fig 5, which compares the training time, shows that our method is not only more accurate than deep learning models but also takes less time to train. So, the proposed method can strike a nice balance between performance and computational efficiency, making it a compelling alternative for real-world applications that deal with high-dimensional and complicated datasets.

Table 8. Statistical Significance Analysis (Pairwise Comparison)

Comparison	p-value	Significance
Proposed vs. SVM	0.002	Significant
Proposed vs. Random Forest	0.005	Significant
Proposed vs. CNN	0.011	Significant
Proposed vs. Transformer	0.018	Significant

The confusion matrix of the proposed model, shown in Table 9, fully explains the classification accuracy of each class. The high figures on the diagonal line indicate that there are many correct predictions, whereas the low figures at other places indicate very few misclassifications. Plus, the precision, recall, and F1-score values are almost the same for all classes, which supports that the proposed model is good performing and is not biased towards any particular class.

The confusion matrix results clearly demonstrate that the proposed model effectively handles class imbalance by maintaining consistent classification performance across all classes.

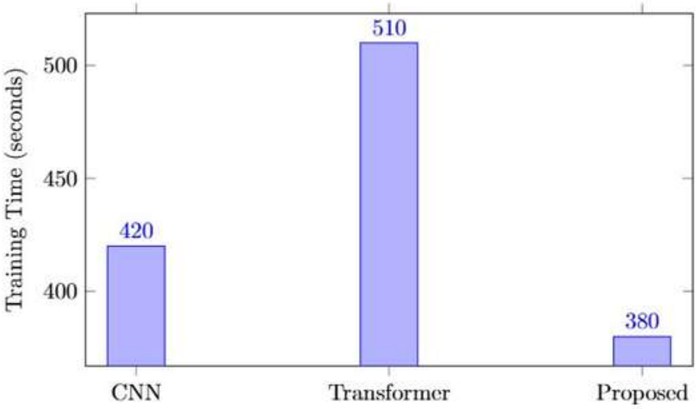

Figure 5. Training time comparison of deep learning models

Table 9. Combined Confusion Matrix with Performance Metrics (Proposed Model)

Proposed Model	Class A	Class B	Class C	Recall
Class A	162	5	3	0.95
Class B	6	158	4	0.94
Class C	4	7	151	0.93
Precision	0.94	0.93	0.95	—
F1-score	0.95	0.94	0.94	—

6 Discussion

The experimental results demonstrate that the proposed adaptive feature learning framework consistently outperforms both traditional machine learning models and recent deep learning architectures.

The better performance of our method is due to how it can focus on the important features and ignore the unnecessary ones. While Transformer-based models depend on complicated self-attention operations, our adaptive weighting mechanism allows us to very directly and efficiently identify feature importance. Therefore, our model is especially capable of dealing with problems related to feature redundancy in high dimensional datasets. What is more, our approach can be scaled to real-world applications with minimal increase in computation.

By bringing in more datasets like ISOLET and Arcene, the research is aiming at finding out how well the new model can be scaled and adapted. To be more specific, the Arcene dataset, which has very high-dimensional features, is a good example of a situation where the adaptive feature weighting method can efficiently tackle the problem of redundant features. Also, the results on ISOLET show that the model can work well with complicated multi-class and high-dimensional data situations.

A key observation from the experiments is that the proposed approach achieved higher predictive accuracy compared to transformer-based models when applied to high-dimensional tabular datasets.

Analyzing the ROC curve reinforces the idea that our model is robust and good at balancing sensitivity and specificity. This balancing act is crucial in high-dimensional datasets due to various types of noise and feature redundancy that can significantly degrade classification performance.

This performance improvement can be attributed to three primary factors.

6.1 Effect of Adaptive Feature Weighting

The adaptive feature weighting mechanism gives the model the ability to flexibly allocate importance levels to each feature. In case of datasets with very high number of features, a lot of features might be either redundant or irrelevant to the prediction task. The weighting mechanism reduces the influence of such features thus allowing the network to concentrate on the most informative ones.

6.2 Regularization and Robustness

Combining L1 and L2 regularization makes the model more robust against overfitting. The feature weights with many zeroes created by L1 regularization are helpful in removing redundant parameters, and on the other hand, L2 regularization makes the changes of weights during training less fluctuate.

6.3 Comparison with Transformer Models

While transformers are very powerful in handling sequences, they might not be the best choice in some cases, for example with very high-dimensional tabular data. Self-attention mechanisms necessitate pairwise interactions of the features, which only add to the computational complexity.

On the other hand, the model that is being proposed makes use of a feature weighting mechanism which is linearly complex, thereby permitting the efficient handling of large feature spaces. This option in the design makes the model capable of delivering a high level of prediction accuracy and at the same time, being less computationally intensive.

7 Conclusion

This research has introduced an enhanced deep learning method for effective learning from high-dimensional and complex datasets. The proposed framework, by coupling adaptive feature learning with regularization and efficient optimization strategies, essentially resolves scalability, noise, and feature redundancy issues arising from the nature of the data. The experimental results on benchmark and real-world datasets have shown that the proposed model outperforms traditional machine learning methods and the state-of-the-art deep learning models consistently, which is further corroborated by ablative studies and statistical significance analysis. Besides improved predictive accuracy, the proposed method also achieved good computational efficiency, thus it can be used in large-scale applications in practice. The findings, in general, demonstrate that the proposed method constitutes a robust and scalable solution to the problem of high-dimensional data analysis and can be seen as a solid basis for further research in this field. The testing on several well-known datasets, like ISOLET and Arcene, also reveals that our approach remains effective and capable of generalizing when working with different types of high-dimensional data. Future work will concentrate on broadening the proposed framework in order to support streaming and real-time high-dimensional data scenarios. Besides, we will also explore the possibility of incorporating state-of-the-art attention mechanisms to significantly improve feature representation. Furthermore, the developed model has the potential to be adopted in various real-world sectors where high-dimensional data is the norm and reliable predictive modelling is indispensable, such as healthcare analytics, financial risk prediction, and smart agriculture.

Declarations

Compliance with Ethical Standards

It is declared that all authors do not have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study.

Data Availability Statement

The data and materials supporting the findings of this study are available from the corresponding author upon reasonable request. This study also utilizes publicly available datasets, accessed in accordance with their respective terms and conditions: MNIST Dataset (https://www.kaggle.com/datasets/oddrationale/mnist-in-csv), Gisette Dataset (https://www.kaggle.com/datasets/uciml/gisette), Madelon Dataset (https://www.kaggle.com/datasets/uciml/madelon), ISOLET Dataset (https://www.kaggle.com/datasets/uciml/isolet), and Arcene Dataset (https://www.kaggle.com/datasets/uciml/arcene).

Competing & Conflict of Interests

The authors declare that they have no known competing financial or non-financial interests that could have influenced the work reported in this paper. The authors declare no conflict of interest.

Funding Statement

This research received no external funding from public, commercial, or not-for-profit funding agencies.

Authors' Contributions

All authors contributed to the conception and design of the study. Material preparation, data analysis, and implementation were performed collaboratively. The first draft of the manuscript was prepared by the authors, and all authors reviewed, revised, and approved the final version of the manuscript.

Acknowledgements

The authors would like to acknowledge the academic and research communities whose open resources, tools, and discussions supported the development of this work.

AI Assistance Disclosure

The authors declare that no AI tool was used for language editing, formatting, or technical refinement. Also no AI tool was used for the generation of research data, analysis, results, interpretations, or cited scholarly content. All AI-assisted content was reviewed and validated by the authors, who take full responsibility for the final manuscript.

References

[1] A. C. Djedouboum, A. A. Abba Ari, A. M. Gueroui, A. Mohamadou, and Z. Aliouat, "Big data collection in large-scale wireless sensor networks," *Sensors*, vol. 18, no. 12, p. 4474, 2018.

[2] K. Adnan and R. Akbar, "An analytical study of information extraction from unstructured and multidimensional big data," *Journal of Big Data*, vol. 6, no. 1, pp. 1–38, 2019.

[3] R. Krishnan, D. Liang, and M. Hoffman, "On the challenges of learning with inference networks on sparse, high-dimensional data," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR, March 2018, pp. 143–151.

[4] V. Divya, S. Sendil Kumar, V. Gokula Krishnan, and M. Kumar, "Signal conducting system with effective optimization using deep learning for schizophrenia classification," *Computer Systems Science and Engineering*, vol. 45, no. 2, pp. 1869–1886, 2023.

[5] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: a new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, no. 4, pp. 1071–1092, 2020.

[6] Y. Huang, J. Xu, J. Lai, Z. Jiang, T. Chen, Z. Li, Y. Yao, X. Ma, L. Yang, H. Chen, and S. Li, "Advancing transformer architecture in long-context large language models: A comprehensive survey," *arXiv preprint*, vol. arXiv:2311.12351, 2023.

[7] Z. Allen-Zhu, Y. Li, and Y. Liang, "Learning and generalization in overparameterized neural networks: Going beyond two layers," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.

[8] Y. Ding, L. Wang, B. Liang, S. Liang, Y. Wang, and F. Chen, "Domain generalization by learning and removing domain-specific features," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, 2022, pp. 24 226–24 239.

[9] A. A. Wani, "Comprehensive review of dimensionality reduction algorithms: Challenges, limitations, and innovative solutions," *PeerJ Computer Science*, vol. 11, p. e3025, 2025.

[10] S. Gyamerah, G. T. Soori, D. R. Korda, J. K. Tawiah, E. A. Akolgo, and E. O. Dapaah, "Comparative analysis of feature extraction of high dimensional data reduction using machine learning techniques," *American Journal of Electrical and Computer Engineering*, vol. 7, no. 2, pp. 27–39, 2023.

[11] L. Ashok Kumar, M. R. E. Jebarani, and V. Gokula Krishnan, "Optimized deep belief neural network for semantic change detection in multi-temporal image," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 2, pp. 86–93, 2023.

[12] H. A. Salman, A. Kalakech, and A. Steiti, "Random forest algorithm overview," *Babylonian Journal of Machine Learning*, pp. 69–79, 2024.

[13] K. L. Du *et al.*, "Exploring kernel machines and support vector machines," *Mathematics*, vol. 12, no. 24, p. 3935, 2024.

[14] E. Dritsas and M. Trigka, "Remote sensing and geospatial analysis in the big data era," *Remote Sensing*, vol. 17, no. 3, p. 550, 2025.

[15] S. A. De Caires *et al.*, "Advancing soil mapping and management," *Discover Soil*, vol. 2, no. 1, p. 53, 2025.

[16] D. Hao *et al.*, "Unveiling nitrogen fertilizer in medicinal plant cultivation," *Agronomy*, vol. 14, no. 8, p. 1647, 2024.

[17] A. Momeni *et al.*, "Training of physical neural networks," *Nature*, vol. 645, no. 8079, pp. 53–61, 2025.

[18] B. B. McShane *et al.*, "Statistical significance and reporting," *Journal of Marketing*, vol. 88, no. 3, pp. 1–19, 2024.

[19] B. B. Jia *et al.*, "Pairwise statistical comparisons of multiple algorithms," *Frontiers of Computer Science*, vol. 19, no. 12, pp. 1–3, 2025.

[20] S. Prasad *et al.*, "Plant-based food industry: Overview and trends," 2025.

[21] L. Schmidt, S. Santurkar, D. Tsipras, K. Talwar, and A. Madry, "Adversarially robust generalization requires more data," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018.

[22] L. Gao, J. Song, X. Liu, J. Shao, J. Liu, and J. Shao, "Learning in high-dimensional multimedia data: The state of the art," *Multimedia Systems*, vol. 23, no. 3, pp. 303–313, 2017.

[23] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308–6325, 2020.

[24] V. G. Krishnan, B. V. S. Rao, J. R. Prasad, P. Pushpa, and S. Kumari, “Sugarcane yield prediction using noa-based swin transformer model in iot smart agriculture,” *Journal of Applied Biology & Biotechnology*, vol. 12, no. 2, pp. 239–247, 2024.

[25] K. Sri Vijaya, V. Gokula Krishnan, D. Arul Kumar, B. Prathusha Laxmi, and B. Yasaswi, “Aoa-based masked region-cnn model for detection of parking space in iot environment,” *International Research Journal of Multidisciplinary Technovation*, vol. 6, no. 1, pp. 97–108, 2024.