CYBERNETICS AND PHYSICS, VOL. 13, NO. 4, 2024, 323–333

NIRSVIT: A NOVEL DEEP LEARNING MODEL FOR MANURE IDENTIFICATION USING NEAR-INFRARED-SPECTROSCOPY AND IMBALANCED DATA HANDLING

Nguyen Van Hieu*
The University of Danang,
ersity of Science and Technology
Vietnam
nvhieuqt@dut.udn.vn

Ngo Le Huy Hien
Leeds Beckett University
United Kingdom
ngolehuyhien@gmail.com

Dinh Minh Toan
The University of Danang,
University of Science and Technology
Viet Nam
toandinh6501@outlook.com

Phan Binh
The University of Danang
Viet Nam
binhphan77373@gmail.com

Phan Minh Nhat
The University of Danang
niversity of Science and Technology
Viet Nam
nhat0299@gmail.com

Phung Thi Anh
The University of Danang
Viet Nam
anhhna17710110@gmail.com

Le Viet Hung
The University of Danang
Viet Nam
hungle.dut@gmail.com

Nguyen Huy Tuong
The University of Danang
Viet Nam
huytuong010101@gmail.com

Article history:
Received 03.08.2024, Accepted 24.12.2024

Abstract

The robust and accurate identification of different forms of manure stands as a pivotal imperative within the domain of agriculture. Near-infrared (NIR) Spectroscopy has emerged as an expeditious, efficient, non-destructive, and reliable approach to addressing this challenging task. NIR spectroscopy has the potential to serve as a valuable tool for the classification and identification of manure varieties. In order to enhance fertilizer identification performance, this study proposes a novel model called NIRsViT which classifies fertilizers by employing a combination of deep learning Vision Transformer model on NIR spectral data. The introduced model's performance outperforms existing deep learning models, with an F1-Score of 86.42% and an accuracy rate of 95.19%. Additionally, the model's classification performance has been significantly improved by proposed imbalanced data processing approaches, Focal Loss, and Upsample, with an F1-Score up to 93.91%, the improved F1-score proved that imbalanced data was considerably solved. The proposed method is a promising approach to handling imbalanced NIR spectral data and acts as a pioneering benchmark for subsequent models in manure identification through NIR spectroscopy. Future research gears toward improving the NIRsViT model's temporal efficiency and computational load, while also testing the introduced imbalanced data handling approach for efficiency comparison across various models and larger datasets.

Key words

NIR Spectroscopy, Deep learning, Imbalanced data, Focal Loss, Vision Transformers

1 Introduction

In current agricultural practices, the provision of essential nutrients to crops is predominantly facilitated through the utilization of fertilizers. The agricultural sector relies on fertilizers to harvest crop yields by furnishing crops with the requisite nutrients for optimal growth and development. Depending on the growth conditions, each type of crop might require different types of fertilizer, therefore, accurate determination of the most suitable fertilizer for a given crop is of paramount importance. Notably, this endeavor holds a formidable challenge, as it necessitates the expeditious categorization and identification of agricultural fertilizers based on their chemical composition, which, in traditional approaches, entails the usage of specialized measurement devices.

However, fertilizer image data is occasionally scarce, and different types of fertilizers sometimes have similar exterior forms, making image-based manure classification a significant difficulty. Image-based manure categorization is one of the most significant obstacles in solving the fine-grained classification problem. Notably, this task is considerably more intricate in comparison to the classification of distinct categories of entities, such as dogs and cats [Wei et al., 2021]. Fertilizer forms share a plethora of physical characteristics, which frequently

pose a notable struggle to the accurate recognition and classification of fertilizers via image processing algorithms. Specifically, conventional image classification models often require an enormous quantity of data and an extensive training period to achieve satisfactory results.

Therefore, Near-infrared (NIRS) spectroscopy has emerged as an effective technique for qualitative analysis of organic substances, especially in food analysis. Compared to the majority of other chemical techniques, NIR technology is swift, reliable, secure, and non-destructive. The process includes recording the molecular vibrations of all molecules having C-H, N-H, O-H, and S-H groups by absorbing near-infrared light (wavelengths ranging from 780nm to 2500nm) at various wavelengths in the sample [Zhang et al., 2022]. Consequently, every resultant spectrum encapsulates the unique characteristics of the corresponding sample. A substantial outcome has been demonstrated in recent research [Zhang et al., 2022] which employs a conventional machine learning model to classify substances using NIR spectra. The Multi-elemental discriminant analysis technique was used to categorize sesame seeds in the research conducted by Young Hee Choi et al. [Choi et al., 2016] with an accuracy of up to 89.4%. SVM and decision tree methods were also utilized in the work of Sylvio Barbon Jr. et al. [Barbon Jr et al., 2018] for chicken flesh identification with a precision score of 77.2%.

Moreover, near-infrared spectroscopy (NIR) has been recently used for classification by combining with analytical methods. The performance of deep learning models in NIR spectral analysis normally outperforms early studies on traditional machine learning techniques, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), K-Nearest Neighbours (KNN), Random Forest, and Support Vector Machine (SVM) [Tan et al., 2022; Huang et al., 2023; Li et al., 2022; Li et al., 2023; Yang et al., 2021]. Deep learning models incorporating NIR spectra have been applied in numerous studies due to this field's exponential growth, and these models consistently yield classification performance of up to 97% precision. Numerous studies have utilized NIR spectra to analyze fertiliser data, as evidenced by recent publications [Tan et al., 2022; Bedin et al., 2021].

Nevertheless, unusual fertiliser types are often seen in practice, leading to data imbalance in classification – an issue that is also commonly encountered in NIR spectral datasets. As a result, models might ignore classes with fewer instances and concentrate primarily on those with more samples. The dataset in Thierry et al.'s research [Hien et al., 2024] on the classification problem using six distinct types of fertilisers underscores this data imbalance issue, with a disproportionately larger quantity of data samples for cattle and poultry manure than for the other fertiliser categories. Although some studies [Morvan et al., 2021; Hong et al., 2019] have mentioned data imbalance remedies; nonetheless, these studies remain facing difficulties in performance, computational complexity, and overfitting.

In recent studies, prominent models such as MLP, CNN, ResNet, and MobileNet have been frequently employed on NIR spectral data for classification tasks [Zhang et al., 2022]. However, emerging evidence indicates that deep learning models based on the recently proposed Vision Transformer architecture by Alexey Dosovitskiy et al. are believed to represent the most sophisticated models for computer vision or image processing tasks [Wang et al., 2020]. With the goal of better understanding the characteristics of NIR data, this study proposes a novel deep learning architecture named NIRsViT, which can be considered as a variation of the Vision Transformers model. Furthermore, the Focal Loss has been used as a Loss function [Khan et al., 2022], which has been confirmed to operate effectively in dealing with unbalanced data. Furthermore, this study consisted of contemplated examining the data balancing method known as Upsample, which involves doubling the data of the classes with fewer samples. The experimental findings of this research point out the ways the proposed NIRsViT model, when coupled with the two Upsample and Focal Loss approaches, substantially improves classification efficiency.

This research has two primary novel contributions:

1. It introduces a new deep learning model called NIRsViT for identifying six different fertiliser types based on NIR spectroscopy.
2. It proposes two novel methods, Upsample and Focal Loss, to rectify the data imbalance issue in NIR spectral data.

The introduced NIRsViT model yields promising results with a 95.19% accuracy rate and an F1-Score of 86.42. Notably, the proposed model performs significantly better when the two proposed methods for handling data imbalance are implemented. While Focal Loss contributes to a 1-2% improvement in F1-Score, a considerable increase in F1-Score, rising to 93.91% is gained by using the Upsample method. The result highlights the unique efficacy of applying the two proposed techniques to address the data imbalance issue, outperforming earlier classification systems.

2 Related Work

2.1 Machine Learning Applications with NIR Spectroscopy

Numerous researchers have recently developed different chemometric methods for NIR spectroscopy to

address the intricacies of NIR spectroscopy-based analysis [Van Huy L, 2023]. Alongside these conventional techniques, the integration of machine learning models within NIR spectroscopy is progressively gaining traction due to their superior efficiency and reliability.

The multi-elemental discriminant analysis machine learning model was used in a recent study by [Choi et al., 2016] to categorise sesame seeds. Regardless, because of the insufficient data on sesame seeds in the data classes, the accuracy of this approach was found to be a modest 89.4%. In another study, [Barbon Jr et al., 2018] classified chicken flesh using traditional machine learning methods such as SVM and Decision Tree. The most effective model, however, only achieved an accuracy of 77.2%, indicating the imperative need for considerable improvement before real-world application becomes feasible. Furthermore, a shortage of data was a challenge experienced by many other researchers [Lin et al., 2017; Mazivila et al., 2020], which intruded on the model's reliability in real-world applications. This occurs because datasets of inadequate diversity engender an incomplete reflection of reality.

Deep learning techniques, as demonstrated by several recent studies [Huang et al., 2023; Yang et al., 2021; Vaswani et al., 2017; Roggo et al., 2007], have outperformed conventional machine learning models. In particular, [Li et al., 2022] proposed a 1-D CNN model for NIR data analysis that outperformed traditional machine learning models such as PLS regression and KSV regression. In [Yang et al., 2021] study, NIR data collected from tea was transformed into images and then applied to CNN, ResNet, and MobileNet. The outcomes illustrated a substantial improvement in performance over traditional machine learning models when applying deep learning to images (pseudo-image). Furthermore, in their pioneering work, [Li et al., 2023] leveraged multiple residual connections to formulate an innovative architecture known as SCNet.

Based on the architecture of the Vision Transformer model, this study introduces a new model named NIRsViT, while concurrently presenting a pseudo-image generated from NIR spectrum data as the input for the model.

2.2 Handling Imbalanced NIR Data

The disparity between classes in a dataset is one of the most pervasive issues in practice, a challenge that also infiltrates the domain of NIR spectral data. Consequently, researchers have sought solutions to address this predicament. In the work of [Hien et al., 2024], machine learning models are re-used to identify anomalous data and concluded that the OC-SVM (One-class SVM) machine learning model is the most effective at dealing with the imbalance in NIR data classification. [Hong et al., 2019] Research is also noteworthy since it employed machine learning models to categorise, utilising a more balanced dataset created by applying the Variational Autoencoder (VAE) network to generate new data [Wong and Yeh, 2024]. However, a substantial quantity of initial data is required to be able to generate data based on ordinary generative models [Zhang et al., 2022]. The object detection problem from [Khan et al., 2022] served as the inspiration for this study to modify Loss Function with Focal Loss to the near-infrared (NIR) spectral classification. The performance of this modification has been outstanding, especially on unbalanced datasets where specific classes have fewer samples [Wen and Furtat, 2023]. Additionally, the Upsample technique is implemented, which is a widely used method for dealing with the problem of data imbalance. By cloning samples repeatedly, this strategy attempts to increase the number of samples of minority classes.

3 Materials and Methods

3.1 Dataset descriptions

The dataset in this research was collected by Thierry Morvan and colleagues in their comprehensive study [Morvan et al., 2021]. This dataset contains data on the near-infrared spectra and chemical makeup of 490 samples of organic fertiliser taken from cattle on 270 farms over two collection campaigns in Brittany, France in 2018 and 2019. The dataset contains the following fertiliser groups: cattle manure, poultry manure, pig manure, poultry droppings, compost, and others. After collecting from the farms, the fertiliser samples underwent a thorough analysis applying analytical techniques approved by the French Standards Organisation (AFNOR) in order to ascertain the concentrations of critical elements such as organic matter, dry matter, N, P, K, Ca, and Mg. The samples were then scanned by a Q-interline AgriQuant B8 near-infrared spectrometer, each sample was scanned three times and the spectral data from 852-2502 nm was saved along with the results of the chemical analysis. Table 1 provides an overview of statistics of the dataset's composition.

The quantity of cattle manure is significantly greater than that of the other types of manure, so the imbalanced data do exist in this dataset. The imbalanced proportion unequivocally demonstrates the imbalance in this dataset, as validated by Wang et al's research [Hong et al., 2019]

3.2 Dataset Preprocessing

The following process is used to smooth and decrease noise in the spectral data:

> Every spectral range in the spectral data undergoes an application of the Savitzky-Golay filter with filter parameters including three polynomial orders and a window size of 21. Through this method, the spectral data is smoothed down, noise is eliminated, continuity is maintained, and it becomes more appropriate for deep learning model training.

Table 1. Sample Size Statistics of fertiliser Groups

Product	Cattle manure	Pig manure	Poultry manure	Poultry droppings	Compost	Other
Sample	276	18	144	27	16	9

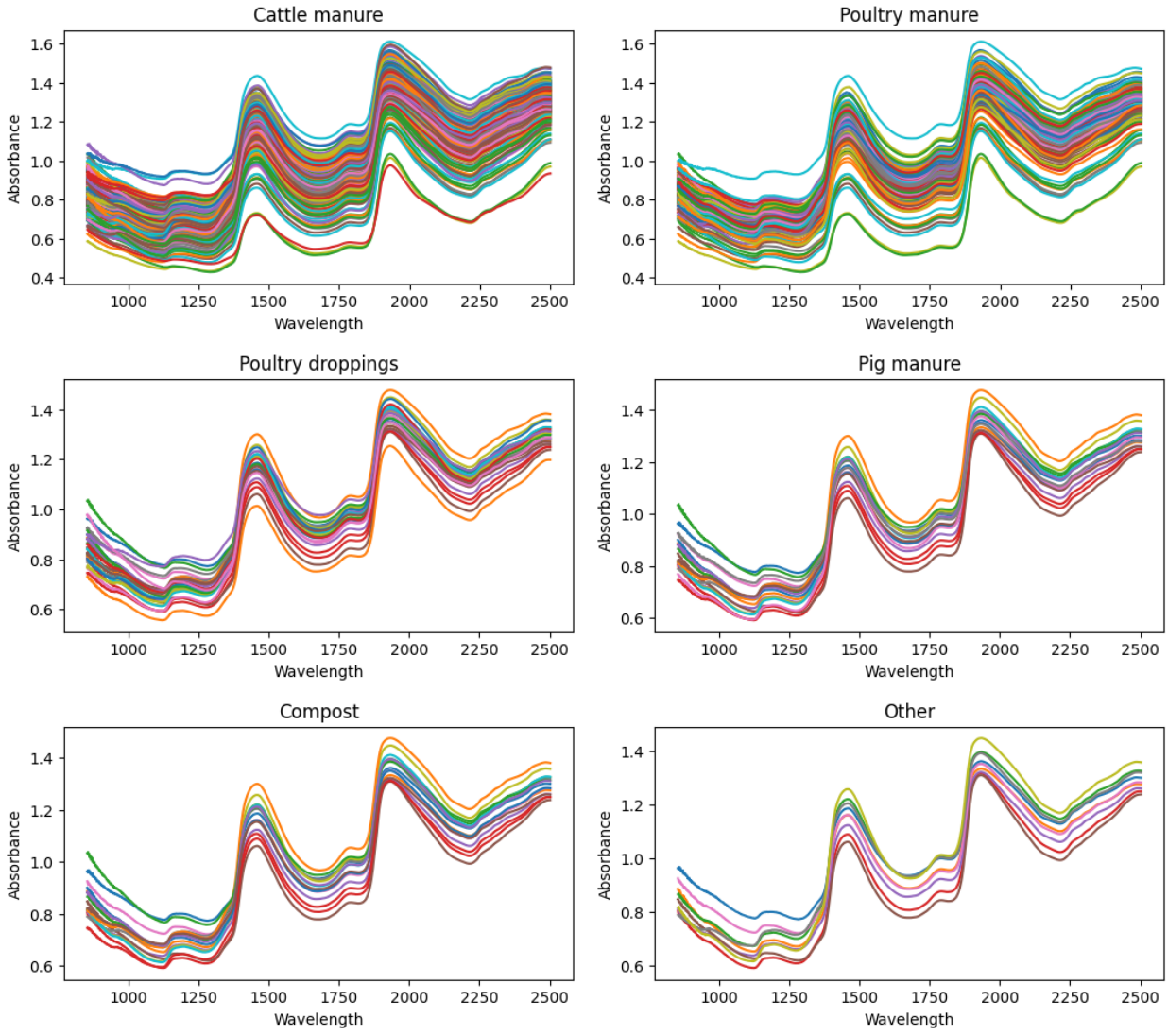

Figure 1. Visualization of NIRS Data of Different Manure Groups

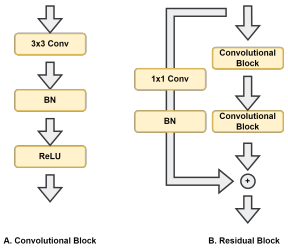

Figure 2. Convolutional Block and Residual Block Architecture

The Standard Normal Variable (SNV) is used for data normalization to bring the dataset to a mean of 0 and a standard deviation of 1 after the spectral data has been smoothed. When it comes to NIR spectral data processing, this is one of the most used normalization techniques.

The combination of the two preprocessing methods, Savitzky-Golay smoothing and Standard Normal Variate (SNV), emerges as the preprocessing method that improves the model most effectively [Zhang et al., 2022].

3.3 The Proposed Model NIRsViT

Convert a one-dimensional spectrum into a two-dimensional image:
Since NIR data is sequence-based, the NIR spectrum data is reshaped into a 22x22x2 image format, shown in Figure 3, to transform into a pseudo-image format, mak-

Table 2. Architecture of MLP, CNN, and ResNet Models

Model	Architecture
MLP	This model consists of fully connected layers with ReLU activation. The first three layers have 512 neurons, while the fourth layer has 256 neurons, and the fifth layer has 128 neurons. The output is a layer of 6 neurons and softmax activation.
CNN	This model includes three consecutive convolutional blocks: 16 kernels 3x3 + BN + ReLU, 32 kernels 3x3 + BN + ReLU, and 64 kernels 3x3 + BN + ReLU. The fourth layer is a global average pooling. The output is a layer of 6 neurons and softmax activation. The architecture of the convolutional block is described in Figure 2 A.
ResNet	This model comprises three consecutive residual blocks: 16 kernels 3x3 + BN + ReLU + Skip connection (kernel 1x1 + BN), 32 kernels 3x3 + BN + ReLU + Skip connection (kernel 1x1 + BN), 64 kernels 3x3 + BN + ReLU + Skip connection (kernel 1x1 + BN). The fourth layer is a global average pooling. The output is a layer with 6 neurons with softmax activation. The architecture of the residual block is depicted in Figure 2 B.

MLP: Multilayer Perceptron, CNN: Convolutional Neural Network, ResNet: Residual Network, BN: Batch Normalization, ReLU: ReLU activation

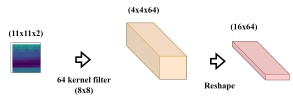

Figure 4. Convolutional Neural Network Architecture

ing it compatible with the Vision Transformer model. The NIR spectrum is expected to be a 968-dimensional vector. We reshape this vector into a tensor with dimensions 22x22x2, resembling an image spectrum.
The architecture of NIRsViT consists of three main components: Convolutional Neural Network, Transformer Encoder, and Classifier Head.

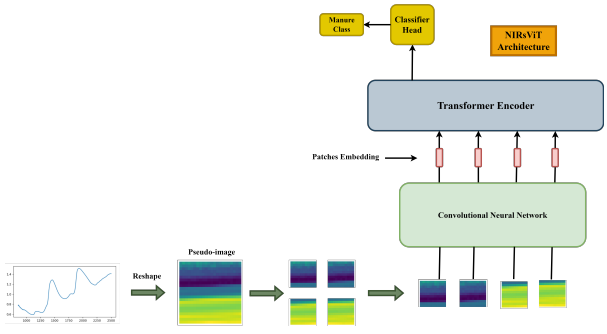

Figure 3. NIRsViT Architecture

The convolutional Neural Network layer should be utilized in place of the original Vision Transformer model's Linear Projection of the flattened patch layer. Employing a dense layer over a flattened patch for image data might not be as effective in extracting features as the Convolutional Neural Network layer (NIR spectrum pseudo-image).

Convolutional Neural Network:

Initially, the pseudo-image data is split up into "patches". Due to the pseudo-image being only 22 × 22 × 2, it is partitioned into 4 "patches", each measuring 11 × 11 × 2, as shown in Figure 3. Following that, as illustrated in Figure 4, each "patch" is sent to the CNN layer to extract the Patch Embeddings with the shape of 16x64. The "red matrix" in Figure 4 is called as Patch Embedding. These Patch Embeddings are then furnished to the Transformer Encoder. Furthermore, the suggested model in this section refrains from using Position Embedding, contrary to the original Vision Transformer architecture. Because using Position Embedding throughout the experiment process not only increases computing cost but also does not significantly improve accuracy.

Transformer Encoder:

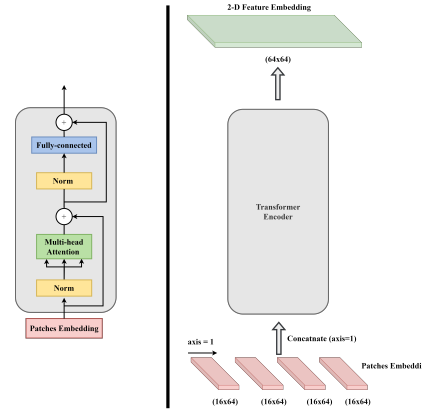

Figure 5. Left: Internal architecture of Transformer Encoder Visualization, Right: Input and Output of Transformer Encoder Visualization

The original Transformer Encoder architecture which was first introduced by [Zhai et al., 2022] is modified with several changes to adapt for the purpose of classifying fertilizers through experimental study rather than deploying a multitude of Transformer Encoder layers, a solitary Transformer Encoder layer is employed. There are two sub-layers in the Transformer Encoder layer [Van Hieu et al., 2023]. One multi-head self-attention layer (with four attention heads) and one normalization layer make up the first sub-layer, while one fully-connected network layer combined with one normalization layer makes up the second [Han et al., 2024]. Furthermore, as can be seen in Figure 5 (Left), a skip connection is applied to every sub-layer. Figure 5 (Right) displays the description of the Transformer Encoder's input and output sizes. In the original Vision Transformer, the authors employ a 'class token' for classification, which might lead to an abundance of other tokens. In our study, we recognize that adding a "class token" could result in an increased computational load. Therefore, we opted not to use a 'class token' in this investigation.

Classifier Head:

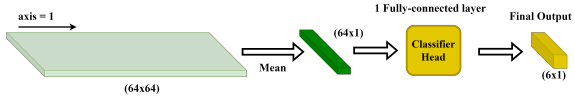

Figure 6. Input and Output of Classifier Head Layer Visualization

A variety of techniques in the Classifier Head layer could be employed depending on the specific classification objective and the dataset's properties. For example, [Dovbnych and Plechawska-Wójcik, 2021] used the Vision Transformer model for classifying forest species and implemented a K-Nearest Neighbours (KNN) model as the classifier that adjusted for the continuous variance of the database's forest species data. Therefore, a single fully connected neural network is employed in this study. This might be attributed to not only the fully connected layer's ability for non-linear separation and its outstanding learning performance on a huge dataset but also its straightforward integration with the Transformer Encoder layer.

The 2D Feature Embedding with a size of 64x64 is extracted by the Transformer Encoder layer. The mean value is then computed on axis = 1, creating a vector with 64 dimensions, shown in Figure 6. After that, the Classifier Head acquires this feature vector, illustrated in Figure 3. The output from the fully connected Classifier Head network layer is represented by the following expression:

$$FC(x) = softmax(Wx + b)$$

Finally, the output of the Classifier Head is a feature vector with 6 dimensions, corresponding to the 6 fertilizer classes.

3.4 Techniques for Overcoming Data Imbalance in the NIRs Spectrum

In the context of a classification task, disparate distribution of training samples among classes can hinder effective learning. This inefficacy arises as the learning process may skip certain classes with limited data samples, and disproportionately focus on learning from classes with more data samples. To address this challenge, this study incorporates two methods to rectify imbalanced data: Upsampling and training the model with the Focal Loss function.

Focal Loss: The loss function implemented in classification tasks, especially in binary classification, commonly is Cross Entropy Loss, which is represented by the following expression:

$$CE(p_t) = -\log(p_t)$$

With:

$$p_t = \begin{cases} p & if \quad y = 1 \\ 1 - p & otherwise \end{cases}$$

While $y = \{\pm 1\}$, which describes class 1 and -1, $p \in [0, 1]$, is the probability of the model's prediction for class $y = 1$. The learning bias that arises from an imbalance in the two-class data is a limitation of the Cross-Entropy Loss function. To attempt to address this problem, a Focal Loss was constructed based on the Binary Entropy Loss formula:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t)$$

Focal Loss is built by adding a coefficient of $(1 - p_t)^{\gamma}$, this coefficient helps to reduce the weight of the majority classes based on the parameter γ. As a result, the model can focus more on the minority classes, thereby significantly improving the accuracy compared to Cross-Entropy Loss. The parameter α_t adjusts the significance of each class in the Focal Loss function.

Upsample: By approximating the larger classes through numerous replications of the small class data samples, the Upsample approach mitigates the challenge of unbalanced data. Compared to the data balancing approach, which generates data as in the work by [Hong et al., 2019], this technique rectifies data imbalance without imposing a great deal of time or computing resources. Moreover, with the goal of preventing overfitting, the Upsample approach is only applied to the training set.

4 Results and Discussion

4.1 Implementation Specifications

Hardware: Within the scope of this study, model training and testing are implemented on CPU Intel Core i5 Coffee Lake – 9300H, GPU Nvidia Geforce GTX 1650 4GB, and RAM 20GB DDR4.

Framework & Environment: All of the source code is written in Python because of its versatility and extensive user community. The models were built using the Pytorch framework on the Windows operating system.

Training Strategy: The models were trained with the same number of epochs = 500. The learning rate was fixed at 0.0001 during training to stabilize the learning process and prevent the learning rate from fluctuating too much during training. The Adam optimizer is employed during training. 2 loss functions, Cross-Entropy Loss and Focal Loss, are utilized for the MLP, CNN, ResNet, and NIRsViT models for the purpose of comparing the performance between the models.

Evaluation Stategy: Traditional machine learning algorithms, including Support Vector Machines (SVM) with linear, polynomial (poly), and radial basis function (RBF) kernels, K-means clustering, Random Forest, XGBoost, and LightGBM, were utilized to benchmark the proposed NIRsViT model. Additionally, deep learning architectures such as Multi-Layer Perceptrons (MLP), Convolutional Neural Networks (CNN), and ResNet were employed to further evaluate model performance. To ensure an objective evaluation of classification outcomes, a K-fold cross-validation [Wong and Yeh, 2019] approach with $k = 5$ was implemented. Additionally, the GridSearchCV method [Jiménez et al., 2008] is used to fine-tune the parameters of the machine learning models, with the details of the tuned parameters described in Table 3.

4.2 Metrics and Evaluation

Metrics: In this study, Accuracy, Precision, and Recall were used as metrics to evaluate the performance of the models. Accuracy determines the ratio between the number of cases predicted correctly and the total number of cases. Accuracy is calculated using the following formula:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

Precision and Recall are used to assess the efficacy of the model's classification quality in classification issues when the class data sets differ considerably. The total number of true positive instances (TP), total number of false cases of negative (FN), and total number of false positive cases (FP) is how a class is defined as positive. The following formulae are used to determine Precision and Recall:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN}$$

F1-Score is a composite metric derived from Precision and Recall to provide a comprehensive measure of classification performance. Since F1-Score combines both accuracy and recall into a single score, it is useful when considering both of these aspects, precluding the prioritization of one at the expense of the other without compromising overall performance. A high F1-Score indicates that the model is able to balance accuracy and recall well. F1-Score is calculated using the following formula:

$$\text{F1-Score} = \frac{\text{2*Precision*Recall}}{\text{Precision + Recall}}$$

Evaluation:

The results in the Table 4 highlight the effectiveness of combining the NIRsViT model with the Focal Loss and Upsampling method to address class imbalance in NIR spectral classification tasks. The NIRsViT model demonstrates superior performance compared to both traditional machine learning and other deep learning methods across all metrics. Among deep learning models using the Cross Entropy Loss function, NIRsViT achieves the highest Accuracy (95.19%), Precision (89.53%), and F1-Score (86.42%). Compared to traditional machine learning models, its Recall and F1-Score outperform even the best-performing model (LightGBM with an F1-Score of 76.06%). This highlights NIRsViT's capacity for effective feature representation and global modeling.

The Upsample method enhances performance across all evaluated models, particularly for imbalanced datasets. Traditional machine learning models benefit significantly, with LightGBM showing an Accuracy increase from 78.57% to 79.13% and an F1-Score rise from 76.06% to 81.18%. For deep learning models, the improvement is even more pronounced. NIRsViT combined with Upsample achieves the highest overall performance: Accuracy (98.10%), Precision (96.78%), Recall (91.21%), and F1-Score (93.91%). This underscores Upsample's effectiveness in mitigating data imbalance and boosting model robustness.

The Focal Loss method proves effective in prioritizing hard-to-classify samples, significantly boosting Precision across all deep learning models. For instance, NIRsViT + Focal Loss achieves a Precision of 92.61%, outperforming its Cross Entropy counterpart (89.53%). However, this gain in Precision comes at the expense of Recall, which decreases slightly due to overemphasis on difficult samples. Despite this trade-off, NIRsViT + Focal Loss achieves a competitive F1-Score of 87.96%, demonstrating its suitability for scenarios prioritizing high Precision.

Although models get high accuracy, these models correctly predict the majority class, such as cattle manure or poultry manure. As observed in Table 5, the models fail to accurately predict the minority class, including pig manure, compost, poultry droppings, and others. Therefore, our imbalanced data handling method

Table 3. Ranges of hyperparameters used in tuning of best results for various machine learning technique

Algorithm	Parameters	Minimum	Maximum	Optimal
SVR (Linear)	cost (C)	9.77×10^{-4}	32	0.01
SVR (Polynomial)	cost (C)	9.77×10^{-4}	32	0.01
	degree	1	5	4
SVR (RBF)	cost (C)	9.77×10^{-4}	32	0.01
	rbf_sigma (sigma)	10^{-10}	0.1	0.05
KMeans	None	None	None	None
Random Forest	tree_depth	1	15	12
	n_trees	1	2000	1850
	min_rows	1	40	32
XGBoost	tree_depth	1	15	12
	n_trees	1	2000	1800
	learn_rate (eta)	0.001	0.4	0.05
	early_stop	3	20	5
LightGBM	tree_depth	1	15	12
	n_trees	1	2000	1800
	learn_rate (eta)	0.001	0.4	0.05
	early_stop	3	20	5

can assist the model in prioritizing the minority class and making accurate predictions for these instances.

4.3 Discussion

The number of samples in this dataset remains suitable for deep learning models, as confirmed by the experimental research presented in Yang et al's research [Yang et al., 2021]. The size of the pseudo-image in NIRS is relatively small, and the NIRsViT model has been customized to be compatible with NIRS data, ensuring that the training process is efficient without imposing a heavy computational load. Additionally, our proposed model's training and testing were conducted on a laptop equipped with an Intel Core i5 Coffee Lake – 9300H CPU, Nvidia GeForce GTX 1650 4GB GPU, and 20GB DDR4 RAM. Spectroscopists can effortlessly implement our proposed model on their computing resources, even with limitations.

Focal loss increases the computational load, leading to longer training times. In contrast, Upsample method doesn't substantially increase the training process, particularly when applied solely to the training data, preventing overfitting. Nonetheless, the Upsample method necessitates manual selection of the upsample frequency for minority classes. On the other hand, focal loss automatically adjusts the weights to reduce emphasis on majority classes and prioritizing minority classes. If you have the best hardware for training, we recommend using the focal loss method. Conversely, if you have limited hardware, we suggest employing the upsample method.

It can be seen that Focal Loss and Upsample consistently yield commendable performance when involving the categorization of unbalanced data, as demonstrated by several prior research. For an unbalanced dataset, it is observed that Focal Loss performs better than Cross-Entropy Loss in terms of performance improvement. When addressing data imbalance effectively and directly by balancing the number of samples in each class, the upsample strategy outperforms Focal Loss in terms of enhancing the performance of deep learning models. The proposed NIRsViT model results in a considerable increase in F1-Score of 7.49. Moreover, when applied to NIR data, models trained on unbalanced data have demonstrated their actual performance.

5 Conclusion

In addressing the challenge of unbalanced data in NIRs data classification tasks, this study proposes the novel NIRsViT model, leveraging the Vision Transformers architecture along with two data upsampling techniques and the incorporation of focal loss. Performance evaluation of NIRsViT against other models, utilizing the unbalanced fertilizer NIRS dataset, is conducted. The

Table 4. Classification Results of Different type of Method using K-fold Cross Validation

Method	Accuracy	Precision	Recall	F1-Score
Machine Learning Method				
SVM (linear)	70.55	66.37	70.55	68.40
SVM (poly)	65.31	58.84	65.31	61.91
SVM (rbf)	56.12	31.49	56.12	40.35
KMeans	39.80	59.52	39.8	47.70
Random Forest	75.51	68.11	75.51	71.62
XGBoost	72.45	66.35	72.45	69.27
LightGBM	78.57	73.71	78.57	76.06
Machine Learning Method using Upsample Method				
SVM (linear)	71.05	71.74	74.72	73.20
SVM (poly)	65.77	63.60	69.17	66.27
SVM (rbf)	56.52	34.04	59.44	43.29
KMeans	40.08	64.34	42.15	50.93
Random Forest	76.05	73.62	79.97	76.67
XGBoost	72.97	71.72	76.73	74.14
LightGBM	79.13	79.24	83.21	81.18
Deep Learning Method using Cross Entropy Loss Function				
MLP	93.57	84.83	82.03	83.41
CNN	94.95	84.99	81.98	83.46
ResNet	95.08	87.00	83.74	85.34
NIRsViT	95.19	89.53	83.52	86.42
Deep Learning Method using Focal Loss Function				
MLP + Focal Loss	94.42	88.67	84.47	86.52
CNN + Focal Loss	95.97	88.35	85.72	87.01
ResNet + Focal Loss	96.55	90.00	81.50	85.52
NIRsViT + Focal Loss	97.41	92.61	83.75	87.96
Deep Learning Method using Upsample Method				
MLP + Upsample	95.96	91.43	86.73	89.01
CNN + Upsample	92.88	88.07	89.99	89.01
ResNet + Upsample	97.67	96.26	85.42	90.51
NIRsViT + Upsample	**98.10**	**96.78**	**91.21**	**93.91**

F1-Score metric is employed for an equitable and unbiased assessment of the proposed approach.

The dual application of Focal Loss during training and Upsample for sample generating significantly contributes to the improved classification performance on imbalanced data. Experimental results, particularly the enhancement in the F1-Score index, affirm the effectiveness of the proposed NIRsViT model in learning from imbalanced data. Research outcomes indicate that the introduced model surpasses the effectiveness of current deep learning models, achieving an impressive F1-Score of 86.42% and an accuracy rate of 95.19%. Moreover, the model's classification performance obtains a

Table 5. This table describes 4 typical samples with true labels and compares the predictions of models using the proposed method with those without the proposed method

Method	**Sample 1 (Pig manure)**	**Sample 2 (Compose)**	**Sample 3 (Poultry droppings)**	**Sample 4 (Other)**
MLP	Cattle manure	Cattle manure	Cattle manure	Cattle manure
CNN	Cattle manure	Cattle manure	Cattle manure	Cattle manure
ResNet	Cattle manure	Poultry manure	Cattle manure	Cattle manure
NIRsViT	Poultry manure	Cattle manure	Cattle manure	Poultry manure
MLP + Focal Loss	Cattle manure	**Compose**	**Poultry droppings**	**Other**
CNN + Focal Loss	Poultry manure	**Compose**	**Poultry droppings**	Compose
ResNet + Focal Loss	**Pig manure**	Cattle manure	**Poultry droppings**	**Other**
NIRsViT + Focal Loss	**Pig manure**	**Compose**	**Poultry droppings**	Pig manure
MLP + Upsample	**Pig manure**	Other	**Poultry droppings**	**Other**
CNN + Upsample	Poultry manure	**Compose**	**Poultry droppings**	Compose
ResNet + Upsample	**Pig manure**	Cattle manure	Other	**Other**
NIRsViT + Upsample	**Pig manure**	**Compose**	**Poultry droppings**	**Other**

substantial enhancement through innovative imbalanced data processing techniques, Focal Loss and Upsample, resulting in an outstanding F1-Score of 93.91%. This proposed method acts as a promising paradigm in addressing imbalanced NIR spectral data, serving as a novel benchmark for future models in the field of manure identification using NIR spectroscopy. Future research endeavours will focus on further enhancements of the proposed NIRsViT model in terms of temporal efficiency and computational load. The introduced approach for handling imbalanced data will be tested to compare and evaluate their efficiency across different models and larger datasets.

6 Acknowledgements

This work was funded by The University of Danang, University of Science and Technology and the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), the code number of Project: VINIF.2023.Ths.119.

References

Barbon Jr, S., Costa Barbon, A. P. A. d., Mantovani, R. G., and Barbin, D. F. (2018). Machine learning applied to near-infrared spectra for chicken meat classification. *Journal of Spectroscopy*, **2018** (1), pp. 8949741.

Bedin, F. C. B., Faust, M. V., Guarneri, G. A., Assmann, T. S., Lafay, C. B. B., Soares, L. F., de Oliveira, P. A. V., and dos Santos-Tonial, L. M. (2021). Nir associated to pls and svm for fast and non-destructive determination of c, n, p, and k contents in poultry litter. *Spectrochimica acta part A: Molecular and biomolecular spectroscopy*, **245**, pp. 118834.

Choi, Y. H., Hong, C. K., Park, G. Y., Kim, C. K., Kim, J. H., Jung, K., and Kwon, J.-H. (2016). A nondestructive approach for discrimination of the origin of sesame seeds using ed-xrf and nir spectrometry with chemometrics. *Food science and Biotechnology*, **25**, pp. 433–438.

Dovbnych, M. and Plechawska-Wójcik, M. (2021). A comparison of conventional and deep learning methods of image classification. *Journal of Computer Sciences Institute*, **21**.

Han, L. N. H., Hien, N. L. H., Van Huy, L., and Van Hieu, N. (2024). A deep learning model for multi-domain mri synthesis using generative adversarial networks. *Informatica*, **35** (2), pp. 283–309.

Hien, N. L. H., Kor, A.-L., Ang, M. C., Rondeau, E., and Georges, J.-P. (2024). Image filtering techniques for object recognition in autonomous vehicles. *Journal of Universal Computer Science*, **30** (1), pp. 49–92.

Hong, X.-Z., Fu, X.-S., Wang, Z.-L., Zhang, L., Yu, X.-P., and Ye, Z.-H. (2019). Tracing geographical origins of teas based on ft-nir spectroscopy: Introduction of model updating and imbalanced data handling approaches. *Journal of analytical methods in chemistry*, **2019** (1), pp. 1537568.

Huang, H.-H., Luo, J.-F., Gan, F., and Hopke, P. K.

(2023). Two revised deep neural networks and their applications in quantitative analysis based on near-infrared spectroscopy. *Applied Sciences*, **13** (14), pp. 8494.

Jiménez, Á. B., Lázaro, J. L., and Dorronsoro, J. R. (2008). Finding optimal model parameters by discrete grid search. In *Innovations in hybrid intelligent systems*, pp. 120–127. Springer.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision: A survey. *ACM computing surveys (CSUR)*, **54** (10s), pp. 1–41.

Li, M., Pan, T., Bai, Y., and Chen, Q. (2022). Development of a calibration model for near infrared spectroscopy using a convolutional neural network. *Journal of Near Infrared Spectroscopy*, **30** (2), pp. 89–96.

Li, Z., Wang, D., Zhu, T., Ni, C., and Zhou, C. (2023). Scnet: A deep learning network framework for analyzing near-infrared spectroscopy using short-cut. *Infrared Physics & Technology*, **132**, pp. 104731.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988.

Mazivila, S. J., Páscoa, R. N., Castro, R. C., Ribeiro, D. S., and Santos, J. L. (2020). Detection of melamine and sucrose as adulterants in milk powder using near-infrared spectroscopy with dd-simca as one-class classifier and mcr-als as a means to provide pure profiles of milk and of both adulterants with forensic evidence: A short communication. *Talanta*, **216**, pp. 120937.

Morvan, T., Gogé, F., Oboyet, T., Carel, O., and Fouad, Y. (2021). A dataset of the chemical composition and near-infrared spectroscopy measurements of raw cattle, poultry and pig manure. *Data in Brief*, **39**, pp. 107475.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., and Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of pharmaceutical and biomedical analysis*, **44** (3), pp. 683–700.

Tan, A., Wang, Y., Zhao, Y., Wang, B., Li, X., and Wang, A. X. (2022). Near infrared spectroscopy quantification based on bi-lstm and transfer learning for new scenarios. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **283**, pp. 121759.

Van Hieu, N., Hien, N. L. H., Van Huy, L., Tuong, N. H., and Thoa, P. T. K. (2023). Plantkvit: A combination model of vision transformer and knn for forest plants classification. *JUCS: Journal of Universal Computer Science*, **29** (9).

Van Huy L, Hien NLH, P. N. H. N. V. (2023). Deep learning model with hierarchical attention mechanism for sentiment classification of vietnamese comments. *Cybernetics and Physics*, **12** (2), pp. 111—20.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, **30**.

Wang, Q., Li, L., Pan, X., and Yang, H. (2020). Classification of imbalanced near-infrared spectroscopy data. In *2020 12th International Conference on Advanced Computational Intelligence (ICACI)*, IEEE, pp. 577–584.

Wei, X.-S., Song, Y.-Z., Mac Aodha, O., Wu, J., Peng, Y., Tang, J., Yang, J., and Belongie, S. (2021). Fine-grained image analysis with deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, **44** (12), pp. 8927–8948.

Wen, X. and Furtat, I. (2023). Nonlinear feedback control based on a coordinate transformation in multi-machine power systems. *Cybernetics and Physics*, **12** (2), pp. 157–161.

Wong, T.-T. and Yeh, P.-Y. (2019). Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, **32** (8), pp. 1586–1594.

Wong, T.-T. and Yeh, P.-Y. (2024). Reliable accuracy estimates from k-fold cross validation. *Cybernetics and Physics*, **13** (3), pp. 228—233.

Yang, J., Wang, J., Lu, G., Fei, S., Yan, T., Zhang, C., Lu, X., Yu, Z., Li, W., and Tang, X. (2021). Teanet: Deep learning on near-infrared spectroscopy (nir) data for the assurance of tea quality. *Computers and Electronics in Agriculture*, **190**, pp. 106431.

Zhai, X., Kolesnikov, A., Houlsby, N., and Beyer, L. (2022). Scaling vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12104–12113.

Zhang, W., Kasun, L. C., Wang, Q. J., Zheng, Y., and Lin, Z. (2022). A review of machine learning for near-infrared spectroscopy. *Sensors*, **22** (24), pp. 9764.