Maize Nitrogen Grading Estimation Method Based on UAV Images and an Improved Shufflenet Network

Weizhong Sun¹,²,³, Bohan Fu¹,²,³ and Zhao Zhang¹,²,³,⁴,*

¹ Key Laboratory of Smart Agriculture System Integration, Ministry of Education, Beijing 100083, China; 2020308130208@cau.edu.cn (W.S.); 2020308130329@cau.edu.cn (B.F.)
² Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture and Rural Affairs, China Agricultural University, Beijing 100083, China
³ College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
⁴ Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo, ND 58102, USA
* Correspondence: zhaozhangcau@cau.edu.cn

Abstract: Maize is a vital crop in China for both food and industry. The nitrogen content plays a crucial role in its growth and yield. Previous researchers have conducted numerous studies on the issue of the nitrogen content in single maize plants from a regression perspective; however, partition management techniques of precision agriculture require plants to be divided by zones and classes. Therefore, in this study, the focus is shifted to the problems of plot classification and graded nitrogen estimation in maize plots performed based on various machine learning and deep learning methods. Firstly, the panoramic unmanned aerial vehicle (UAV) images of maize farmland are collected by UAV and preprocessed to obtain UAV images of each maize plot to construct the required datasets. The dataset includes three classes—low nitrogen, medium nitrogen, and high nitrogen, with 154, 94, and 46 sets of UAV images, respectively, in each class. The training set accounts for eighty percent of the entire dataset and the test set accounts for the other twenty percent. Then, the dataset is used to train models based on machine learning and convolutional neural network algorithms and subsequently the models are evaluated. Comparisons are made between five machine learning classifiers and four convolutional neural networks to assess their respective performances, followed by a separate assessment of the most optimal machine learning classifier and convolutional neural networks. Finally, the ShuffleNet network is enhanced by incorporating SENet and improving the kernel size of the Depthwise separable convolution. The findings demonstrate that the enhanced ShuffleNet network has the highest performance; its classification accuracy, precision, recall, and F1 scores were 96.8%, 97.0%, 97.1%, and 97.0%, respectively. The RegNet, the optimal model among deep learning models, achieved accuracy, precision, recall, and F1 scores of 96.4%, 96.9%, 96.5%, and 96.6%, respectively. In comparison, logistic regression, the optimal model among the machine learning classifiers, attained accuracy of 77.6%, precision of 79.5%, recall of 77.6%, and an F1 score of 72.6%. Notably, the logistic regression exhibited significant enhancements of 19.2% in accuracy, 17.5% in precision, 19.5% in recall, and 24.4% in the F1 score. In contrast, RegNet demonstrated modest improvements of 0.4% in accuracy, 0.1% in precision, 0.6% in recall, and 0.4% in the F1 score. Moreover, ShuffleNet-improvement boasted a substantially lower loss rate of 0.117, which was 0.039 lower than that of RegNet (0.156). The results indicated the significance of ShuffleNet-improvement in the nitrogen classification of maize plots, providing strong support for agricultural zoning management and precise fertilization.

Keywords: UAV images; machine learning; deep learning; data augmentation; maize nitrogen detection

1. Introduction

Maize is an important food crop and industrial raw material in China, and maintaining stable maize yields plays a vital role in national food security [1]. The nitrogen content
is an important factor affecting maize growth, and insufficient nitrogen can significantly impact the number of grains per spike [2]. Within a certain range of fertilizer applications, the number of grains per spike increases with the amount of nitrogen fertilizer applied. However, excessive nitrogen fertilization has little effect on the number of grains per spike [3]. Excessive fertilizer application can cause various problems, including increased costs, wasted resources, plant lodging, and environmental pollution. Therefore, nitrogen estimation for maize during the growing period is the basis for proper nitrogen fertilizer application, which helps improve maize yield and fertilizer utilization rates and avoid soil, air, and water pollution caused by blind fertilizer application. This means that the technology can be applied to crop growth monitoring and nitrogen fertilizer management. Hence, it has significant economic, social, and ecological benefits.

The traditional method for estimating the nitrogen content of maize is to indirectly grasp the nitrogen status of crop leaves through the soil and plant analyzer development (SPAD) measurements of the SPAD-502 chlorophyll meter. Many research results have been found on applying chlorophyll meters for nitrogen deficits and nitrogen requirement prediction, crop growth evaluation, and water and fertilizer management measures in rice, maize, sorghum, and spinach crops [4–6]. Studies have shown that the SPAD values of chlorophyll meter readings for different fertility periods can indirectly reflect the chlorophyll content of crop leaves and the total plant nitrogen content, and can further guide the follow-up application of nitrogen fertilizer. However, for large farmland areas, the method of chlorophyll meter determination consumes a great deal of workforce and material resources and may result in some drawbacks. Therefore, there is an urgent need to develop a high-throughput real-time nitrogen estimation method for agricultural fields.

With the rapid development of information technology and the remarkable improvement of the agricultural information level, the methods for estimating crop nitrogen contents based on remote sensing technology have gradually emerged in recent years. Vigneau et al. used a tractor carrying a HySpex VNIR1600-160 (Norsk Elektro Optikk, Norway) hyperspectral camera to scan and obtain spectral data from 400 to 1000 nm in the wheat canopy to establish a quantitative model to estimate the canopy leaf nitrogen content with a coefficient of determination (R2) of 0.889 [7]. Tao et al. used a power exponential relationship model. Additionally, they achieved good predictions of nitrogen content in wheat leaves with a correlation coefficient of 0.67, which reached a highly significant level [8]. The development of UAV-based remote sensing systems has taken remote sensing and precision agriculture further. The use of UAV for crop monitoring offers great possibilities to obtain field data in a simple, fast, and cost-effective way compared to previous methods [9]. Liu et al. used a UHD185 hyperspectral spectrometer (450–950 nm) carried by UAV to obtain hyperspectral images of wheat at the joining stage, heading stage, flowering stage, and grain filling stage, and used the sensitive bands obtained via a correlation analysis to establish a multiple regression model and B.P. neural network model, which could better estimate canopy leaf nitrogen contents. The model’s R2 value reached 0.948 [10]. In addition to UAVs, the use of starboard data is gradually appearing on our horizon. Delloye et al. used Sentinel-2 and SPOT satellite data to estimate the canopy nitrogen threshold via the inversion of wheat canopy chlorophyll contents using artificial neural networks and other algorithms, which provided a scientific basis for rapid decision-making based on crop nitrogen fertilization requirements [11].

From the above research results, it can be seen that the data type for nitrogen detection has gradually shifted from mainly relying on hyperspectral to multispectral data. The platforms for spectral acquisition have mainly moved to near-ground, airborne, and satellite-based with the development of remote sensing technology. With the rapid development of artificial intelligence technology in recent years, smart agriculture offers an effective solution to today’s agricultural sustainability challenges. For crop selection, crop management, and crop production prediction, artificial intelligence technologies have far-reaching implications [12]. The methods for nitrogen detection have also evolved from linear models to more complex mathematical models such as partial least squares.
regression (PLSR), support vector machines (SVM), back-propagation neural networks, and genetic algorithms (G.A.) for the modeling and estimation of the nitrogen content in recent years. However, the prevailing methods for estimating crop nitrogen contents tend to obtain the specific nitrogen content of a plant precisely. There is a relative gap in the study of plot-based nitrogen grading tasks. Precision agriculture has recently been a hot field of agricultural science research. The core of precision agriculture is the precise management of soil nutrients, and zoning management is the main means to achieve this management. Specifically, zoning management manages areas with similar production potential, similar nutrient utilization, and similar environmental effects as a management unit [13]. The fertilizer dosage should be adjusted according to the soil nutrient status and crop nutrient demands of different management units to improve the soil production potential, improve the nutrient utilization rate, reduce environmental pollution, and improve the crop yield and quality [14]. Scientific and reasonable management zoning can guide farmers’ field water and fertilizer management and provide an economical and effective means for accurately managing farmland nutrients. Therefore, the perspective of maize nitrogen estimation was changed to the problem of plot classification for the purpose that it could meet the needs of zoning management in precision agriculture.

In recent years, deep learning has made significant contributions in the field of agriculture as well. Gulzar presented a fruit image classification model that leveraged deep transfer learning and the MobileNetV2 architecture. The model enabled efficient and accurate fruit classification by utilizing pretrained deep learning models in the agricultural domain [15]. Mamat et al. improved the image annotation technique for fruit classification [16]. By employing deep learning algorithms, this method achieved more precise identification and annotation of agricultural product images, enhancing the effectiveness of the fruit classification. Aggarwal et al. employed a stacked ensemble approach based on artificial intelligence to predict protein subcellular localization in confocal microscopy images [17]. This method offered highly accurate predictions for protein localization, serving as a valuable tool for protein analyses in agriculture research. Dhiman et al. presented a comprehensive review of image acquisition, preprocessing, and classification methods for detecting citrus fruit diseases [18]. They summarized image capture, preprocessing, and classification techniques, providing important guidance for disease detection and management in agriculture through an in-depth analysis of the existing literature. In addition, Ünal et al. provided an overview of smart agriculture practices specifically in potato production, which contributed to the development of intelligent solutions for potato production [19].

In this research, machine learning (ML) and deep learning (DL) methods were used to train the training set based on plot-specific UAV images, save the model, and test it on the test set to output the nitrogen ratings of corn plots. The performances of the ML and DL methods were compared on the test set, and the relatively optimal ShuffleNet network model was improved by considering the large farmland-embedded device application scenario. The improved ShuffleNet network model performs well and allows the double guarantee of accuracy and efficiency for the nitrogen grading problem in maize plots.

The main contributions of this paper are as follows:

- Proposing methods based on the most recent advances in machine learning and deep learning approaches, as these methods are proven to be remarkably accurate and effective. They can be advantageously utilized in the field of crop management support through innovative precision agriculture approaches;
- The perspective on maize nitrogen estimation is shifted to the problem of plot classification to better align with the requirements of zoning management in precision agriculture.

The rest of the paper is organized as follows. Section 2 introduces the collection and preparation process for the dataset in the paper, the selection of the experimental methods, and the model improvements. Section 3 presents a comparison and analy-
ysis of the experimental results. Finally, Section 4 concludes with a summary of the assessment’s findings.

2. Materials and Methods

UAVs represent a low-cost alternative inspection and data analysis technology mainly used for monitoring and spraying in precision agriculture. The maturing artificial intelligence technology also has far-reaching implications for crop management and detection in precision agriculture. In this study, the two techniques were combined to collect maize images from maize test plots and the ML and DL models were adopted to derive the results for maize nitrogen levels in each plot. The specific flow chart is shown in Figure 1. First, the UAV images of maize farmland were acquired by UAV. Afterwards, we began constructing the dataset and proceeded to select the model. The ML dataset was generated via the feature extraction of UAV images, actual measurements of maize SPAD values, and labeling, and the DL dataset was generated using RGB images from UAV images, which corresponded to the labels one by one. The ML and DL methods were selected for classifying the maize nitrogen levels. Finally, a performance evaluation was performed for comparison and improvement between models.

![Flowchart of maize nitrogen level division using UAV images and different ML and DL algorithms.](image)

**Figure 1.** Flowchart of maize nitrogen level division using UAV images and different ML and DL algorithms.

2.1. Data Collection

The data for the study were obtained from the experimental maize field from the Jilin Academy of Agricultural Sciences in Gongzhuling, Jilin Province, China. The experimental maize field is located at 124°82’ E, 43°52’ N, and the altitude of the experimental field is 207 m above sea level. Plots were used as the minimum counting units to meet the needs of land zoning management. The trial field was divided into three large areas: the common hybrid trial area (7 rows and 96 columns, totaling 180 plots), the high-nitrogen hybrid trial area (3 rows and 60 columns, totaling 45 plots), and the high-nitrogen self-incompatible line trial area (6 rows and 60 columns, totaling 120 plots). Each area of the trial field was set up with one protection row at the top and bottom and two protection columns at the left and right. The experiment started on 8 May 2022, with 100 pounds of compound fertilizer applied per acre during maize growth, and the images were collected after 3 months of growth.

The Genie4RTK UAV, manufactured by DJI Innovation Technology, Shenzhen, China, as depicted in Figure 2a, was employed for intermittent aerial photography at an altitude
of 30 m on 25 August 2022. Throughout the UAV’s flight, it remained perpendicular to the ground, ensuring minimal image distortion when capturing images of various plots. The weather conditions on the day of image acquisition were sunny, with light winds. The UAV operation commenced at 12:14, precisely during the peak solar elevation of the day, facilitating a uniform distribution of sunlight across the experimental field. Consequently, during the vertical image acquisition process, no shadows were observed across the different plots. Additionally, the UAV’s preset flight path was set at a 31° angle relative to the experimental field to facilitate the creation of a UAV panoramic view of the cornfield using DJI Terra software. DJI Terra was used for stitching to generate six panoramas of UAV images of the test field (containing RGB images and five remote sensing parameters stored in TIFF format, including the LCI, GNDVI, NDRE, NDVI, and OSAVI). Among them, the RGB image panorama of the test field is shown in Figure 2b.

Figure 2. (a) Physical picture of the Sprite4RTK UAV, (b) Panoramic view of an RGB image of the experimental field.

To obtain the true label of maize nitrogen, the actual content of maize nitrogen is also needed. The SPAD value is the relative value of the leaf chlorophyll content, and studies have shown that the leaf SPAD values are significantly and positively correlated with the total nitrogen content at different fertility stages, which represents that the SPAD value can be used to reflect the actual nitrogen content and nitrogen nutrition status of the crop [20]. A SPAD-502 chlorophyll meter was used to detect the SPAD values of each column of maize, and then we derived the mean value of the detected SPAD values of maize in the plot to derive the average SPAD, which represents the nitrogen status of the plot.

A total of 294 plots were selected for nitrogen measurements, consisting of 135 plots for the common hybrid trial area, 45 plots for the high-nitrogen hybrid trial area, and 114 plots for the high-nitrogen self-incompatible line trial area. Each plot was treated as an experimental subject, resulting in a total of 294 experimental subjects for this experiment. To capture the necessary data, we captured UAV images of each experimental subject by cropping the panorama of the experimental field. Additionally, we measured the leaf SPAD of each maize plant within the plot to obtain the average SPAD values for each experimental subject.

2.2. Image Preprocessing

In the real world, the images collected are often incomplete, inconsistent, and highly susceptible to noise (in this paper, our team mainly solve the problem of data inconsistency). Image preprocessing is necessary to facilitate the analysis and improve the accuracy [21].
The UAV panorama of the test field was generated after reading the file by DJI Terra. To obtain the UAV images of each cell in the test field, it is necessary to perform image segmentation on the panorama of the test field. Since each category within the UAV image is the same size, the RGB image was used as the base for image segmentation. The same segmentation method was then applied to the other categories to complete all segmentation steps. The specific RGB image segmentation steps are as follows:

1. **Angle correction:** Since the experimental maize field is tilted 31° to the left in the RGB panorama, an artificial angle correction is performed to square the experimental maize field in the visual field by rotating the RGB panorama clockwise by 31°;

2. **Grayscale:** The RGB image is a three-channel image, including R, G, and B channels. To make the subsequent line segmentation easier, we begin by converting the RGB image to grayscale using Formula (1):

   \[
   Gray = 0.3R + 0.59G + 0.11B
   \]

   (1)

3. **Remove protected rows and columns:** Protected rows and columns (mentioned in Section 2.1) are not our experimental objects. Hence, cropping the RGB panorama manually by removing them is required;

4. **Row and column segmentation to obtain RGB images of individual plots:** Since the number of planting rows and columns is known, and most of them are uniformly planted, the lengths of individual plots can be calculated directly according to the relationship between the number of rows and columns and the lengths of images. However, due to overlapping and coverage of the plants in the segmentation process, artificial fine-tuning is implemented. When extending a row of plants to other rows, we slightly increase the width of the corresponding image area while decreasing the width of the adjacent area. Similarly, when extending a column of plants to other columns, we slightly increase the length of the corresponding image area while decreasing the length of the adjacent area. The final RGB images of each block are obtained, with each individual block image having a size of 145 × 355 pixels. Due to occlusion in some images, the resolution of each block’s RGB image may slightly fluctuate up or down. These images are stored in TIFF format, and an example figure is shown in Figure 3.

![Figure 3. An example image of a single maize plot RGB image.](image)

To obtain the UAV images of each cell, the segmentation of the RGB images was used to guide the creation of the remaining category images. Here, 294 × 6 images were obtained, which were stored in TIFF format.

2.3. **Data Preparation**

2.3.1. **Machine Learning Dataset Preparation**

Since only the average SPAD per corn plot was obtained, the data needed to be converted into categorical information. Therefore, a K-means clustering analysis based on the obtained average SPAD of each plot was performed after removing outliers [22]. Our team set \( K = 3 \) (divided into three intervals of low, medium, and high nitrogen) to obtain three clustering centers and the category labels of each maize plot, and the result of the clustering analysis is shown in Figure 4. The ranges of nitrogen content (PNC) in our maize...
plots are 0.8–1.5% for low nitrogen, 1.5–2.5% for medium nitrogen, and over 2.5% for high nitrogen. The numbers for the low, medium, and high nitrogen categories in the dataset are 154, 94, and 46, respectively. Finally, each UAV image is matched to its corresponding label to obtain the required dataset for this type of ML.

![K-means analysis for maize SPAD value](image)

**Figure 4.** Result of the K-means analysis of maize SPAD values, where the dashed lines are the classification boundaries of the three categories calculated from the three clustering centers.

2.3.2. Deep Learning Dataset Preparation

Due to the limited volume of our dataset, which consists of only 294 RGB images, it failed to meet the high demands of neural networks for extensive data. Additionally, there was an imbalance between classes. Data augmentation on the segmented RGB image dataset was performed to address these challenges. Data augmentation techniques can reduce the risk of overfitting effectively and improve the accuracy and robustness of DL models [23]. In this experiment, enhancement techniques such as rotation, mirroring, Gaussian noise, luminance, and Gaussian blur were selected to expand the DL dataset from 294 to 1933 images, of which 640, 721, and 552 were low-, medium-, and high-nitrogen images, respectively. Each RGB image was matched with the corresponding label obtained in Section 2.3.1, resulting in the dataset needed for our DL model.

2.4. Machine Learning Methods

2.4.1. Feature Extraction

For the ML method, feature extraction is an important process that involves two aspects. First, there is the statistical analysis and transformation of images to extract the required features. Second, there is the transformation and operation of group measurements of a pattern to highlight its representative features [24]. For this paper, feature extraction involves extracting, combining, and transforming the channel parameters to generate a new feature subset based on the ML dataset. Figure 5 shows the flow of the feature extraction and Table 1 presents the features selected in this paper.

Traditional nitrogen detection methodologies often relied on single-feature analyses, which lacked the capacity to capture the multidimensional intricacies inherent in the datasets. To overcome this limitation, our research focuses on an integrated approach, where RGB, spectral indices, and texture features are synergistically employed to unearth novel insights into nitrogen detection.

RGB data derived from images captured by remote sensing devices present distinct advantages for nitrogen detection [25]. RGB data inherently encapsulate information about
leaf coloration, which is influenced by the chlorophyll content. By incorporating RGB features, our model effectively captures the spatial distribution and structural variations within the vegetation, enabling the better differentiation of nitrogen-rich and nitrogen-deficient regions.

Figure 5. Diagram of the feature extraction process.

Spectral indices are widely recognized for their sensitivity to specific biochemical and biophysical properties of vegetation [26]. Indices such as the normalized difference vegetation index (NDVI) serve as proxies for vegetation health and photosynthetic activity, both closely associated with nitrogen availability. Leveraging spectral indices in our feature selection process affords the ability to gauge plant vitality and consequently infer the nitrogen content.

Texture features, rooted in the spatial arrangement and distribution of pixel intensities, offer valuable supplementary information for nitrogen detection [27]. Texture descriptors, including but not limited to Haralick features, enable the characterization of fine-grained patterns and structures within vegetation, thereby capturing subtle variations indicative of nitrogen levels. Integrating texture features augments the discriminative capacity of our model, furnishing it with the ability to discern intricate variations that might elude single-feature analyses.

2.4.2. Data Preprocessing

Before feeding the features into the classifier, the data must be preprocessed to account for variations in feature magnitudes. Specifically, the data need to be converted from various specifications or distributions into a standardized format called dimensionless data. The goal is to accelerate the solution, enhance the model’s precision, and prevent a specific feature with an unusually wide range of values from disproportionately affecting distance calculations. Data normalization was used for data preprocessing. The equation is shown in (2).

$$x^* = \frac{x - \mu}{\sigma}$$  \hspace{1cm} (2)

where $\mu$ is the mean of the current feature and $\sigma$ is the variance of the current feature.

2.4.3. Support Vector Machine

A support vector machine (SVM) is a powerful machine learning algorithm for solving classification problems [29]. It separates different classes of samples by finding a dividing hyperplane in the sample space, while maximizing the minimum distance from two points sets to this hyperplane. In two-point sets, the edge points that are nearest to the hyperplane are called support vectors. SVMs can be categorized into linear SVMs and nonlinear SVMs. Linear SVM is suitable for dealing with linear problems but not nonlinear problems. For nonlinear problems, kernel functions can transform them from a low-dimensional space to a high-dimensional space, which can be treated as linear problems. The commonly used kernel functions are the linear kernel function, polynomial kernel function, Gaussian kernel function, Laplace kernel function, and Sigmoid kernel function. The exceptional
performance of the support vector machine in handling small sample, nonlinear, and high-dimensional datasets is the reason for selecting it as a classifier.

Table 1. Description and formulation of color and texture features of UAV images.

<table>
<thead>
<tr>
<th>Number</th>
<th>Feature</th>
<th>Description</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R</td>
<td>Red channel</td>
<td>R_{ij}</td>
</tr>
<tr>
<td>2</td>
<td>G</td>
<td>Green channel</td>
<td>G_{ij}</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>Blue channel</td>
<td>B_{ij}</td>
</tr>
<tr>
<td>4</td>
<td>NRI</td>
<td>Normalized red index</td>
<td>R/(R + G + B)</td>
</tr>
<tr>
<td>5</td>
<td>NGI</td>
<td>Normalized green index</td>
<td>G/(R + G + B)</td>
</tr>
<tr>
<td>6</td>
<td>NBI</td>
<td>Normalized blue index</td>
<td>B/(R + G + B)</td>
</tr>
<tr>
<td>7</td>
<td>R/(R + G − B)</td>
<td>/</td>
<td>R/(R + G + B)</td>
</tr>
<tr>
<td>8</td>
<td>G/(R + G − B)</td>
<td>/</td>
<td>G/(R + G + B)</td>
</tr>
<tr>
<td>9</td>
<td>B/(R + G − B)</td>
<td>/</td>
<td>B/(R + G + B)</td>
</tr>
<tr>
<td>10</td>
<td>NDI</td>
<td>Normalized difference index</td>
<td>128 \times \left(\frac{G-R}{R+G}\right) - 1</td>
</tr>
<tr>
<td>11</td>
<td>ExG</td>
<td>Excess green index</td>
<td>2G - R - B</td>
</tr>
<tr>
<td>12</td>
<td>ExR</td>
<td>Excess red index</td>
<td>1.3R - G</td>
</tr>
<tr>
<td>13</td>
<td>CIVE</td>
<td>The color index of vegetation extraction</td>
<td>0.441R - 0.811G</td>
</tr>
<tr>
<td>14</td>
<td>ExGR</td>
<td>Excess green minus excess red index</td>
<td>ExG - ExR</td>
</tr>
<tr>
<td>15</td>
<td>NGRDI</td>
<td>Normalized green minus difference index</td>
<td>(G - R)/(G + R)</td>
</tr>
<tr>
<td>16</td>
<td>VEG</td>
<td>Vegetative index</td>
<td>G/(R^{0.667} - B^{0.333})</td>
</tr>
<tr>
<td>17</td>
<td>NEG</td>
<td>Normalized excessive green index</td>
<td>2.8G - R - B</td>
</tr>
<tr>
<td>18</td>
<td>MExG</td>
<td>Modified excess green index</td>
<td>1.262G - 0.884R - 0.311B</td>
</tr>
<tr>
<td>19</td>
<td>GMR</td>
<td>Green minus red</td>
<td>G - R</td>
</tr>
<tr>
<td>20</td>
<td>H</td>
<td>Hue channel</td>
<td>\frac{\mu_{\text{H}}}{V_{\text{H}}}</td>
</tr>
<tr>
<td>21</td>
<td>S</td>
<td>Saturation channel</td>
<td>\frac{\mu_{\text{S}}}{V_{\text{S}}}</td>
</tr>
<tr>
<td>22</td>
<td>V</td>
<td>Value channel</td>
<td>\frac{\mu_{\text{V}}}{V_{\text{V}}}</td>
</tr>
<tr>
<td>23-74</td>
<td>GLCM</td>
<td>52 features</td>
<td>details in the paper [28]</td>
</tr>
<tr>
<td>75</td>
<td>NDVI</td>
<td>Normalized difference vegetation index</td>
<td>(NIR + R)/(NIR - R)</td>
</tr>
<tr>
<td>76</td>
<td>GNDVI</td>
<td>Green light normalized differential vegetation index</td>
<td>(NIR - G)/(NIR + G)</td>
</tr>
<tr>
<td>77</td>
<td>NDRE</td>
<td>Normalized differential red edge index</td>
<td>(NIR - RED EDGE)/(NIR + RED EDGE)</td>
</tr>
<tr>
<td>78</td>
<td>LCI</td>
<td>Land cover index</td>
<td>\frac{(N^*)^{2}((T_{\text{max}}-T_{0})^2)}{(T-T_{0})}</td>
</tr>
<tr>
<td>79</td>
<td>OSAVI</td>
<td>Optimized soil-adjusted vegetation index</td>
<td>(NIR - R)/(NIR + R + 0.16)</td>
</tr>
</tbody>
</table>

2.4.4. K-Nearest Neighbor

The K-nearest neighbor (KNN) algorithm is based on the minority following the majority. It has a remarkable prediction effect, is resilient to outliers, and finds extensive use in diverse application areas. It has been applied to maize pest detection with good results [30]. The KNN algorithm operates by comparing the features of test data with those in a known training set. Specifically, when the labels for the training set are known, the algorithm finds the K most similar data points in the training set. It assigns the category corresponding to the test data as the most frequent classification among these K points. The learnable hyperparameter is the number of neighbors K. The KNN algorithm works well with small and large amounts of low-dimensional data, consistent with the mini-batch dataset. The KNN algorithm was selected as a classifier.

2.4.5. Decision Tree

A decision tree (DT) classifies data by a set of rules [31]. It provides a rule-like approach to which values will be obtained under which conditions. There are two types of DTs: the classification tree and regression tree. The classification tree generates DTs for discrete variables, and the regression tree generates DT for continuous variables. The generation
process of DT consists of three main steps: feature selection, DT generation, and pruning. DTs are suitable for both numerical and nominal purposes, where the outcome of a variable takes values from a finite set of targets. They can allow data collection and extract the rules embedded in certain columns. A DT was chosen as a classifier because the discrete features extracted in this study are applicable in its context.

2.4.6. Random Forest

A random forest (RF) algorithm is an integrated algorithm [32]. First, it randomly selects different features and training samples to generate many decision trees. Then, it integrates the results of these decision trees to perform the final classification. The RF approach is widely used in real analyses. Compared to decision trees, the RF approach shows a significant improvement in accuracy and enhances the robustness of the model, thereby reducing its susceptibility to attacks. An RF algorithm was selected as one of the classifiers because of its superior accuracy compared to other algorithms.

2.4.7. Logistic Regression

A logistic regression (LR) algorithm is an algorithm that applies linear regression to classification problems. It involves adding individual attribute features by weighted summation to obtain the combined information, converting the linear model fit values to label probabilities using a sigmoid function and obtaining the optimal coefficients by minimizing the cross-entropy cost function. The LR model is simple, and the output values are probabilistically meaningful for the linear correlation features constructed in this study. This is precisely why LR was selected to be used as the classifier.

2.5. Deep Learning Methods

DL is a method for learning the intrinsic patterns of sample data based on artificial neural networks, which learn the abstract representations of data hierarchically, allowing for the automatic processing of complex tasks [33]. Unlike the traditional ML method, the main advantage of DL is that it can automatically learn features from data without the need for the human design of the features. In addition, the DL models are highly flexible and scalable to cope with various complex tasks and data types. We experimented using DL to rank the nitrogen levels and tested its feasibility with the classical network, AlexNet. Since the target application for large farmlands involves mobile or embedded devices that require high efficiency and accuracy, three lightweight networks including RegNet, ShuffleNet, and EfficientNet, were employed to achieve high accuracy and fast detection. Finally, model improvement was performed on the well-performing ShuffleNet. The implementation details will be provided in the following sections.

2.5.1. AlexNet

AlexNet is a classical convolutional neural network model in the field of DL, which won the ImageNet image recognition competition in 2012. AlexNet mainly consists of five convolutional layers and three fully connected layers, in which the ReLU function is used for the activation function and maximum pooling is used for the pooling operation. Using GPU-accelerated training, AlexNet achieved significant performance improvements, and its Top-5 error rate was reduced to 15.3% [34]. The AlexNet network structure is shown in Figure 6.

2.5.2. RegNet

RegNet is an image classification model based on a neural network. The network structure is designed through a novel approach—the new network design paradigm, which combines the benefits of a manual network design and neural network search. The figure depicting its network structure is shown in Figure 7. RegNet’s design takes inspiration from the trade-off between the network depth and width. A traditional network design approaches usually improve the performance of a model by increasing the network
depth or width. However, this results in higher computational complexity for the model. RegNet accomplishes the objective of enhancing the model performance without adding to the computational load by dynamically adjusting the network depth and width, as demonstrated in [35].

Figure 6. AlexNet network architecture.

Figure 7. RegNet structure diagram: (a) RegNet network framework diagram; (b) body module structure diagram; (c) stage / module structure diagram; (d) block x structure diagram when stride = 1; (e) block x structure diagram when stride = 2.

2.5.3. EfficientNet

As shown in Figure 8, EfficientNet is a convolutional neural network structure that aims to balance the computational efficiency and accuracy. It uses compound scaling, which means the model is efficient by reducing the network’s depth, width, and resolution. The authors further improved the model’s performance during training by searching for the best combination of hyperparameters using the auto-ML method. In addition, EfficientNet uses a new group convolution method, MB convolution, which combines depth-separable convolution and residual connectivity and can improve the expressiveness and generalization performance of the model. Due to its efficiency and portability, EfficientNet is widely used in mobile devices and embedded systems [36].

2.5.4. ShuffleNet

ShuffleNet is an efficient convolutional neural network structure mainly used for image classification and target detection tasks [37]. ShuffleNet is characterized by a reduced number of parameters and computational complexity while maintaining its accuracy. This feature is achieved by employing two key techniques: group convolution and channel
shuffle. Figure 9 illustrates the group convolution process used to reduce the computational cost of convolving different feature maps of the input layer. Specifically, the feature maps are partitioned into groups, each involving a separate set of kernels. This grouping strategy enables parallel processing and parameter sharing, significantly reducing the computation and memory requirements.

As shown in Figure 10, group convolution degrades the performance of the model when all of the input feature map information needs to be considered. For this problem, a channel shuffle module is added between two group convolutions to disrupt the channel order, thereby enabling the exchange of information between different groups.

For ShuffleNet-V2, the authors presented four guidance summaries: balancing the number of groups by utilizing group convolution, avoiding network fragmentation, and reducing element-level operations. The authors analyzed the shortcomings of the ShuffleNet-V1 design and incorporated improvements to create ShuffleNet-V2 as outlined in the guidance summary. The overall structure of ShuffleNet-V2 is shown in Figure 11a and the module (stage) structure diagram of ShuffleNet-V2 is shown in Figure 11b,c. The main improvement introduced in the V2 version is the channel split technique, which involves dividing the input feature map into two parts in the channel dimension. The left branch of the split is mapped equally, while the right branch contains three successive convolutions with identical input and output channels, conforming to guidance 1. Furthermore, in the
V2 version, the previously used group convolution for $1 \times 1$ convolution is replaced with ordinary convolution, aligning with guidance 2.

![Figure 10](image)

**Figure 10.** Schematic diagram of the channel shuffle module.

![Figure 11](image)

**Figure 11.** (a) The overall structure of ShuffleNet-V2. (b) Stage structure when stride = 1. (c) Stage structure when stride = 2.

2.5.5. ShuffleNet-Improvement

The model accuracy and model efficiency are a pair of contradictory indicators. ShuffleNet has achieved a very high level of performance in terms of model efficiency. Regardless, the model accuracy is a little behind, and our improvement goal is to improve the model accuracy by sacrificing some of the model efficiency to meet the engineering requirements of large-scale farmland nitrogen detection. The specific improvements are reflected in the following two aspects:
1. The computational distribution of ShuffleNet-V2 is shown in Table 2. Upon examining the table, our team see that the DW convolution contributes a relatively small amount of the computation costs, while the majority of the computation is concentrated on the $1 \times 1$ convolution. Therefore, all $3 \times 3$ DW convolutions were replaced with $5 \times 5$ DW convolutions. This will not increase the computational weight too much but will also improve the model accuracy.

2. In deep convolutional neural networks, selecting the feature channels is crucial to improve the performance. However, traditional neural networks often do not consider the correlation between channels when processing the weights of feature channels, thereby failing to utilize the information between feature channels fully. To address this problem, Hu et al. presented a channel attention mechanism based on the squeeze-and-excitation (SE) module, called SENet, in 2018 [39].

<table>
<thead>
<tr>
<th>Layer (Type)</th>
<th>Flops</th>
<th>Params</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv2D</td>
<td>59.3 M</td>
<td>1.4 K</td>
</tr>
<tr>
<td>Group Conv2D (stage2)</td>
<td>3.0 M</td>
<td>57.1 K</td>
</tr>
<tr>
<td>Group Conv2D (stage3)</td>
<td>5.3 M</td>
<td>218.2 K</td>
</tr>
<tr>
<td>Group Conv2D (stage4)</td>
<td>4.3 M</td>
<td>756.2 K</td>
</tr>
<tr>
<td>Group Conv2D (stage5)</td>
<td>3.6 M</td>
<td>2.5 M</td>
</tr>
<tr>
<td>AvgPool2D</td>
<td>4.4 K</td>
<td>0</td>
</tr>
<tr>
<td>Fully Connected (F.C.)</td>
<td>515.1 K</td>
<td>51.2 K</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>75.3 M</td>
<td>3.6 M</td>
</tr>
</tbody>
</table>

The channel attention mechanism of SENet adaptively learns the importance of each channel, thereby strengthening the important feature channels in the network and suppressing the insignificant ones. As shown in Figure 12, in SENet, the global average pooling layer is first performed using the squeeze operation to obtain the global average of each channel. Afterward, the excitation operation is performed using two fully connected layers that learn the weights of each channel by mapping the c-weight coefficients. These weights are finally multiplied by each channel in the input feature map to obtain the feature map adjusted by the channel attention mechanism.

![Figure 12. Schematic diagram of SENet.](image)

The channel attention mechanism was borrowed from SENet and applied to ShuffleNet, and the size of the DW convolution was boosted. The overall structure of the improved network is shown in Figure 13.

![Figure 13. Schematic diagram of the ShuffleNet-V2-improvement network structure.](image)
2.6. Model Evaluation Metrics

In classification problems, evaluation metrics can help us assess the performance and effectiveness of a model. In classification problems, four metrics, namely the accuracy (ACC), precision (P), recall (R), and F1 score (F1), are commonly used to assess the accuracy of the model [40]. Their formulas are described as shown in (3)–(6):

\[
\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}
\]

\[
P = \frac{TP}{TP + FN} \tag{4}
\]

\[
R = \frac{TP}{TP + FN} \tag{5}
\]

\[
F1 = \frac{2 \times P \times R}{P + R} \tag{6}
\]

The true positive (T.P.) is the number of successfully predicted positive samples. The true negative (T.N.) is the number of examples successfully predicted as negative. False positives (F.P.s) represent the number of negative samples incorrectly predicted to be positive. The false negative (F.N.) is the number of positive samples incorrectly predicted as negative.

However, this study deals with a multi-classification problem that requires consideration of the predictive performance of the three categories of low, medium, and high nitrogen for the precision (P), recall (R), and F1 score. While recalibration is not necessary for accuracy (ACC), it is crucial to evaluate the overall functionality of this identification system. Considering the impact of the incomplete balance of the data volume on the results, the weighted-average method was used to describe its performance in terms of the Weighted-P, Weighted-R, and Weighted-F1, whose formulas are described in (7)–(10) as follows [41,42]:

\[
\text{Weighted} - P = \sum_{i=1}^{3} P_i \times W_i \tag{7}
\]

\[
\text{Weighted} - R = \sum_{i=1}^{3} R_i \times W_i \tag{8}
\]

\[
\text{Weighted} - F1 = \sum_{i=1}^{3} F1_i \times W_i \tag{9}
\]

\[
W_1 : W_2 : \cdots : W_n = N_1 : N_2 : \cdots : N_n \tag{10}
\]

where W represents the category’s weight and the actual number of samples in the category.

Since the precision and recall are equally important to us, the Weighted-F1 is selected as the more balanced and comprehensive metric to measure the accuracy of our model. In addition to the above metrics for evaluating the accuracy, the frames per second (FPS) and model size are also used to evaluate the model’s efficiency. Using Python’s time function, we recorded the time just before reading the image and after outputting the model. To determine the program’s running time, here we calculate the difference between these two moments, repeat the process 100 times to obtain an average, and finally calculate the frames per second (FPS) by taking the inverse of the running time.

3. Results

3.1. Machine Learning Results

Since testing the model on a dataset divided in a specific way may not generalize well to new data, a ten-fold cross-validation method is employed. This involves randomly
dividing the dataset into ten equal parts, using nine for training and one for testing, repeating this process ten times, and averaging the results. The obtained results are shown below.

Figure 14 shows the performances of different ML classifiers for this dataset. Upon observing the histogram, it becomes apparent that the SVM performs well in terms of accuracy, with values up to 79%. However, LR surpasses the SVM in terms of the accuracy, recall, and F1 score, with a negligible difference in accuracy. The P, R, and F1 score were considered more representative indicators than ACC since they account for the number of instances in each category. Based on these metrics, it is concluded that the LR classifier outperforms the other models for this dataset.

In addition to the classifier’s accuracy, the classifier, efficiency of the classifier, and model size are also important factors that affect the model’s performance. Figure 15 compares the FPS and model size results of five ML classifiers. The FPS comparison shows that the five classifiers have similar performance levels. After analyzing the data, it can be found that the small difference in FPS (ranging from 39.1 images per second to 40.0 images per second) is because reading the image features accounts for the majority (about 90%) of the running time weight, while the time spent on feature operations by the classifiers is relatively small. Therefore, the FPS levels between different classifiers are not significantly different. In real production applications, this small difference can be ignored. By observing the model size comparison, it can be found that the SVM has the largest model size due to its kernel function and relatively complex computation costs, while the rest of the models are at the level of 4–5 kb in size.

![Figure 14. Accuracy results for 5 ML classifiers (SVM, KNN, DT, LR, RF) with ten-fold cross-validation rules.](image-url)
3.2. Deep Learning Results

3.2.1. Deep Learning Model Accuracy Comparison

Here, 80% of the dataset was used as the training set and the remaining 20% as the test set. We performed training for 100 epochs, and the results of the test set are shown in Figure 16. The figure shows that the enhanced ShuffleNet network performs well in all four evaluation metrics, namely the accuracy, precision, recall, and F1 score, while achieving the lowest test loss value. This indicates that introducing the attention mechanism and increasing the convolutional kernel size of the DW convolution effectively improve the model accuracy.

Figure 15. (a) Computational efficiency comparison of 5 ML classifiers (SVM, KNN, DT, LR, RF). (b) Model size comparison of 5 ML classifiers (SVM, KNN, DT, LR, RF).

Figure 16. Cont.
3.2.2. Deep Learning Running Efficiency and Model Size Comparison

According to Figure 17a, when measured by FPS, Alexnet exhibits the highest detection efficiency, capable of detecting 491 images per second. AlexNet was the first convolutional neural network created with relatively shallow layers and naturally fast operation speeds. ShuffleNet has the second-best performance with 98 images/s. The improved ShuffleNet only experiences a slight decrease in FPS to 90 images/s.

As shown in Figure 17b, comparing the five model sizes, ShuffleNet uses its lightweight advantage, with a model size of only 4.95 MB. Despite the improvement, the model size is only 0.92 MB larger than other network model sizes. ShuffleNet-improvement is still way ahead in model size.

4. Discussion
4.1. Recommendation for Machine Learning Classifiers and Deep Learning Models

Combining the performances of the five classifiers in terms of accuracy and efficiency, LR was selected as the best ML classifier for the maize nitrogen ranking problem due to its higher accuracy and smaller model size.

Considering the model accuracy, efficiency, and size, ShuffleNet-improvement has the highest model accuracy, which is sufficient for the nitrogen classification of large farmland.
areas. Regarding the model efficiency, the 91 images/s detection speed is not as fast as AlexNet and ShuffleNet but it is sufficient to meet the engineering needs. In terms of the model size, ShuffleNet-improvement also inherits ShuffleNet's small model size, and the model size of 5.87 MB is ideal in embedded devices. It can be said that ShuffleNet-improvement optimizes the performance of ShuffleNet while maintaining its lightweight design, resulting in an optimal model for practical engineering applications. Therefore, ShuffleNet-improvement is recommended as the best model for DL.

4.2. Comparison of Machine Learning Models and Deep Learning Approaches

This section is based on the results of LR, the optimal ML model, and ShuffleNet-improvement, the optimal DL model. ShuffleNet-improvement achieves the highest level of accuracy in the rank division. For ShuffleNet-improvement, the accuracy, precision, recall rate, and F1 score are 96.8%, 97.0%, 97.1%, and 97%, respectively. For LR, the accuracy, precision, recall rate, and F1 score are 77.6%, 79.4%, 77.6%, and 72.6%, respectively. Compared to LR, the accuracy, precision, recall rate, and F1 score are increased by 19.2%, 17.6%, 19.5%, and 24.4%, respectively. Regarding the model efficiency, ShuffleNet-improvement achieves 91 FPS while LR only manages 39 FPS. This puts ShuffleNet-improvement ahead in terms of performance. The model sizes for ShuffleNet-improvement and LR were 5.87 MB and 0.19 MB, respectively. While the models trained by ShuffleNet-improvement are significantly larger than those of LR at 5.87 MB, they still have a small footprint for embedded systems in field applications. This is because the software sizes in such applications typically measure several gigabytes. ShuffleNet-improvement is the optimal model for maize nitrogen level classification based on UAV images.

Romualdo et al. studied maize nitrogen nutrition diagnoses using artificial vision techniques, and the global classification accuracy was 94% for the optimal classifier model test for leaves grown in four different N-concentration culture environments [43]. In comparison, our optimal classifier not only has a higher classification accuracy rate (96.8%) but also does not have the limitation of the detection time for the leaf-specific growth period. In addition, our evaluation system is more robust for the multiclassification problem of nitrogen classification. By calculating the precision, recall, and F1 score, the problem of unobjective evaluation metrics caused by unbalanced sample sizes among classes is avoided.

4.3. Research Limitations and Future Research Directions

This study aimed to detect and grade the nitrogen levels in large farmland areas to guide crop production. However, there are certain objective conditions that impose constraints on the feasibility of our experimental protocol. Two research limitations can be identified as follows:

(1) The distance between the UAV and the crop may have an impact on the accuracy of the model. Therefore, it is recommended to maintain a consistent height of 30m, which was used during the data collection, in the actual application process;

(2) Since the light intensity can impact the shooting effect of the UAV and the light intensity in the actual application scene is highly likely to differ from that in our experimental setup, it becomes a crucial factor that limits the accuracy of our model. As a result, it is recommended to validate the model using field datasets collected under varying light conditions in the actual application scenario.

In future research, there are two extension directions as follows:

(1) Higher data collection heights result in higher efficiency but inevitably lower accuracy. Therefore, it would be beneficial to explore the impact of height on maize nitrogen grading. A balance between collection efficiency and accuracy can be found to give the optimal data collection height;

(2) This research is based on static images, and in the future the purpose is to embed our algorithms into UAVs for the real-time detection of nitrogen levels in large agricultural
fields. This will be achieved through target detection, thereby contributing to the development of unmanned farms.

5. Conclusions

In this paper, UAV images of corn fields were first collected and preprocessed. The ML and DL methods were tested based on the UAV image dataset to demonstrate the applicability of computer vision technology for maize nitrogen grading classification. Among the five ML classifiers (SVM, KNN, DT, LR, RF), LR was selected as the best ML classifier for the maize nitrogen classification problem due to its higher accuracy and smaller model size. Among the DL algorithms, several models were tested, and it was found that ShuffleNet outperformed AlexNet, EfficientNet, and RegNet. Therefore, it was improved based on ShuffleNet by introducing the SENet attention function and improving the DW convolution kernel size to generate a new ShuffleNet-improvement model. Finally, our preferred choice for top ML and DL algorithms was the ShuffleNet-improvement, which stood out as our preferred choice due to its exceptional accuracy (96.8%), precision (97%), recall (97.1%), F1 score (97%), high FPS (91 images/s), and relatively small model size (5.87 MB). Future research studies could explore the effect of height on maize nitrogen grading and develop an embedded system for real-time practical applications in conjunction with target detection technology. This study contributes to the further development of precision agriculture and provides a strong support for the management of nitrogen fertilization during maize growth.

Author Contributions: Conceptualization, W.S. and Z.Z.; methodology, Z.Z.; software, W.S.; validation, W.S., B.F. and Z.Z.; formal analysis, W.S.; investigation, W.S.; resources, Z.Z.; data curation, W.S.; writing—original draft preparation, W.S.; writing—review and editing, B.F. and Z.Z.; visualization, W.S.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by Chinese Universities Scientific Fund (funding no. 15053343).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

6. Yin, W.; Chai, Q.; Guo, Y.; Fan, H.; Yu, A. The physiological and ecological traits of strip management with straw and plastic film to increase grain yield of intercropping wheat and maize in arid conditions. Field Crops Res. 2021, 271, 108242. [CrossRef]


15. Gulzar, Y. Fruit image classification model based on MobileNetV2 with deep transfer learning technique. Sustainability 2023, 15, 1906. [CrossRef]

16. Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing image annotation technique of fruit classification using a deep learning approach. Sustainability 2023, 15, 901. [CrossRef]


18. Dhiman, P.; Kaur, A.; Balasaraswathi, V.R.; Gulzar, Y.; Alwan, A.A.; Hamid, Y. Image acquisition, preprocessing and classification of citrus fruit diseases: A systematic literature review. Sustainability 2023, 15, 9643. [CrossRef]


23. Shorten, C.; Khoshgoftar, T.M. A survey on image data augmentation for deep learning. J. Big Data 2019, 6, 60. [CrossRef]


31. Wang, Y.; Shi, W.J.; Wen, T.Y. Prediction of winter wheat yield and dry matter in North China plain using machine learning algorithms for optimal water and nitrogen application. Agric. Water Manag. 2023, 277, 108140. [CrossRef]

32. Wei, X.Y.; Huang, X.; Bo, L.Y.; Mao, X.M. Vertical distribution of nitrogen content in spring maize leaves and its hyperspectral inversion under different film mulching and irrigation levels in northwest arid region. J. China Agric. Univ. 2022, 27, 13–21. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.