Machine Learning Research
2025, Vol. 10, No. 1, pp. 44-52
https://doi.org/10.11648/j.mlr.20251001.15

Research Article

Red Onion Seed Quality Classification Using Transfer Learning Approaches

Tarekegn Walle Yirdaw[1], Ermias Melku Tadesse[2, *], Endalkachew Hiwote[3], Abebaw Mebrate[3], Ambaw Mulatu[4]

[1]Information System Department, Kombolcha Institute of Technology, Wollo University, Kombolcha, Ethiopia
[2]Department of Information Technology, Kombolcha Institute of Technology, Wollo University, Kombolcha, Ethiopia
[3]Software Engineering Department, Kombolcha Institute of Technology, Wollo University, Kombolcha, Ethiopia
[4]Department of Computer Science, Institute of Technology, Wollo University, Kombolcha, Ethiopia

Abstract

An essential vegetable that is grown all over the world and eaten in a variety of ways is the onion (Allium cepa L.). A common condiment used to improve food flavor is onion. Around the world, red onion seed (A. fistulosum) is cultivated in a variety of temperate and tropical settings. It is grown in China and Japan, among other places, worldwide. A. fistulosum is grown across Ethiopia in various regions. In 2012, 3,281,574 tons of output were obtained from 30,478 hectares of coverage. Allium fistulosum covers the Amhara area over 8000 hectors, which is 26% of our country. For export, red onion seed is separated based on quality. Red onion seed quality separation or categorization is essential to the trade process. It aids in making people marketable. In Ethiopia, this procedure is carried out manually, which has a number of drawbacks like being less effective, inconsistent, and prone to subjectivity. To address this problem, we use pre-trained transfer learning model VGG, GoogleNet, and ResNet50 for quality classification of red onion seed. The main procedures include image preprocessing, resizing, data augmentation, and prediction. The model trained on 470 datasets collected from different agricultural fields in south Gondar libo kemkem and fogera woreda. We use various augmentation strategies to expand the dataset. Ten percent of the dataset was set aside for testing, ten percent for validation, and eighty percent for training. For VGG19, VGG16, GoogleNet, and ResNet, the model's classification accuracy for the input image is 99%, 100%, 100%, and 86%, respectively.

Keywords

Allium Fitsulosum, Red Onion Seed, Visual Geometry Group, GoogleNet, ResNet, Pretrained Models, Transfer Learning

1. Introduction

All civilizations began with agriculture. The development of productivity is the main objective [1]. On Earth, agriculture is most important for human sustenance. Many farming practices and equipment have been digitized throughout the years to ensure faster production of higher quality. The rising demand in the farming sector necessitates the employment of a productive process to cultivate agricultural products [2].

An extremely important vegetable that is grown all

*Corresponding author: ermiasmelku3400@gmail.com (Ermias Melku Tadesse)

Received: 10 March 2025; **Accepted:** 2 April 2025; **Published:** 29 April 2025

throughout the world and utilized in many different recipes is the onion, or Allium cepa L. [3]. The onion (Allium cepa L.), sometimes known as the common onion or bulb onion, is the only member of the Alliaceous family and the most widely cultivated species of the genus Allium. It is grown in more than 130 countries worldwide. China and India are the world's largest producers of onions, with the US, the Netherlands, Egypt, and Iran coming in second and third, respectively [4].

Red onion seed is grown throughout Ethiopia in various regions. In 2012, 3,281,574 tons of output were obtained from 30,478 hectares of coverage. Allium fistulosum covers the Amhara region above 8000 hectors, which is 26% of our country. The seeds of A. fistulosum were sold domestically and exported to other nations [5].

Although there are studies on the detection of onion diseases, there are few studies on the classification of red onion seed production quality. According to experts in agriculture, the marketing quality of seed production is low. In order to categorize red onion seed production quality into three group's namely healthy, foreign objects, and immature/shriveled grains our study use transfer learning techniques.

2. Related Work

By gathering 1432 photos and classifying them into 14 kinds, the paper Automatic and Fast Classification of Barley Grains from Images suggests a deep learning method for barley classification. The paper has done noise removal and augmentation preprocessing techniques. The data is divided in to 80% training, 10% validation and 10% testing. They used transfer learning VGG 16 for feature extraction and classification and they achieved an overall accuracy of 94% [6].

The work Onion agronomy and post-harvest handling Manual describes we can get good quality onion seed production by putting the seed on water using floatation to get quality seed [3]. But this is a traditional system and it takes time as well as needs more manpower. It also needs drying the seed immediately and strainer must be used.

The study, "Participatory Evaluation and Demonstration of Onion Spacing in Irrigated Agriculture at Kencho Kebele in Uba Debre Tsehay Woreda, Southern Ethiopia," examines how to maximize onion yields by applying 416.56mm irrigation water at five-day intervals and utilizing four different onion plant spacing levels: 8 cm, 10 cm, 12 cm, and farmer's practice or planting with own practice. The Red Creole onion variety has been the subject of the work. With a score of 17.178 t/ha, the research concludes that a better onion plant spacing of 10 cm is recommended [7]. Although the effort concentrates on the onion's bulb yield, we also need to improve the quality of the seed yield.

In the Jimma zone at the Agaro Agricultural Research Sub-center, an adaptation trial of three improved onion varieties (Nan thus, Adama red, and Bombe red) with one local check was conducted using RCBD design for two consecutive seasons in order to evaluate the adaptability and yield performance of onion (Allium cepa L.) varieties in Southwestern Ethiopia. Lastly, with an overall yield performance of 71.51 tons/ha, the Bombe red variety proved to be well-adapted and produced the maximum yield in both cropping seasons [8]. However, since the investigation focusses on bulb yield performance, we must evaluate the quality of onion seed production.

3. Research Methodology

3.1. Proposed System

The system architecture, concepts and algorithms used, evaluation metrics and their representation in relation to the research, and software tools and libraries used to carry out the implementation are all covered in detail in this chapter.

3.1.1. System Architecture

The architecture for putting the suggested system into practice is shown in the accompanying diagram.

3.1.2. Data Collection and Dataset Preparation

Getting pictures of the onion seed was the first stage. Images gathered from South Gondar's agricultural areas with the assistance of the specialists identified in the table are included in this study. Using tools like color and weight, experts mark or label the image according to their class. In all, 470 images of the onion seed were obtained. Every image was in JPEG format and RGB color. Three categories were applied to the images: foreign stuff, immature shrivel led grain, and healthy. A training set and a testing set were created from the classified data. The neural network was trained using the training set, and the system was tested using the test set. Training accounted for 70% of the data, testing for 15%, and validation for 15%. The photos were trained and validated using three transfer learning models: VGG, Google Net, and ResNet 50. the professionals who worked on our dataset's annotation.

3.1.3. Data Augmentation for Proposed Model

The study uses augmentation to boost the quantity of photos. To add to the data, we use an image data generator. Additionally, the preprocessed data images are enhanced.

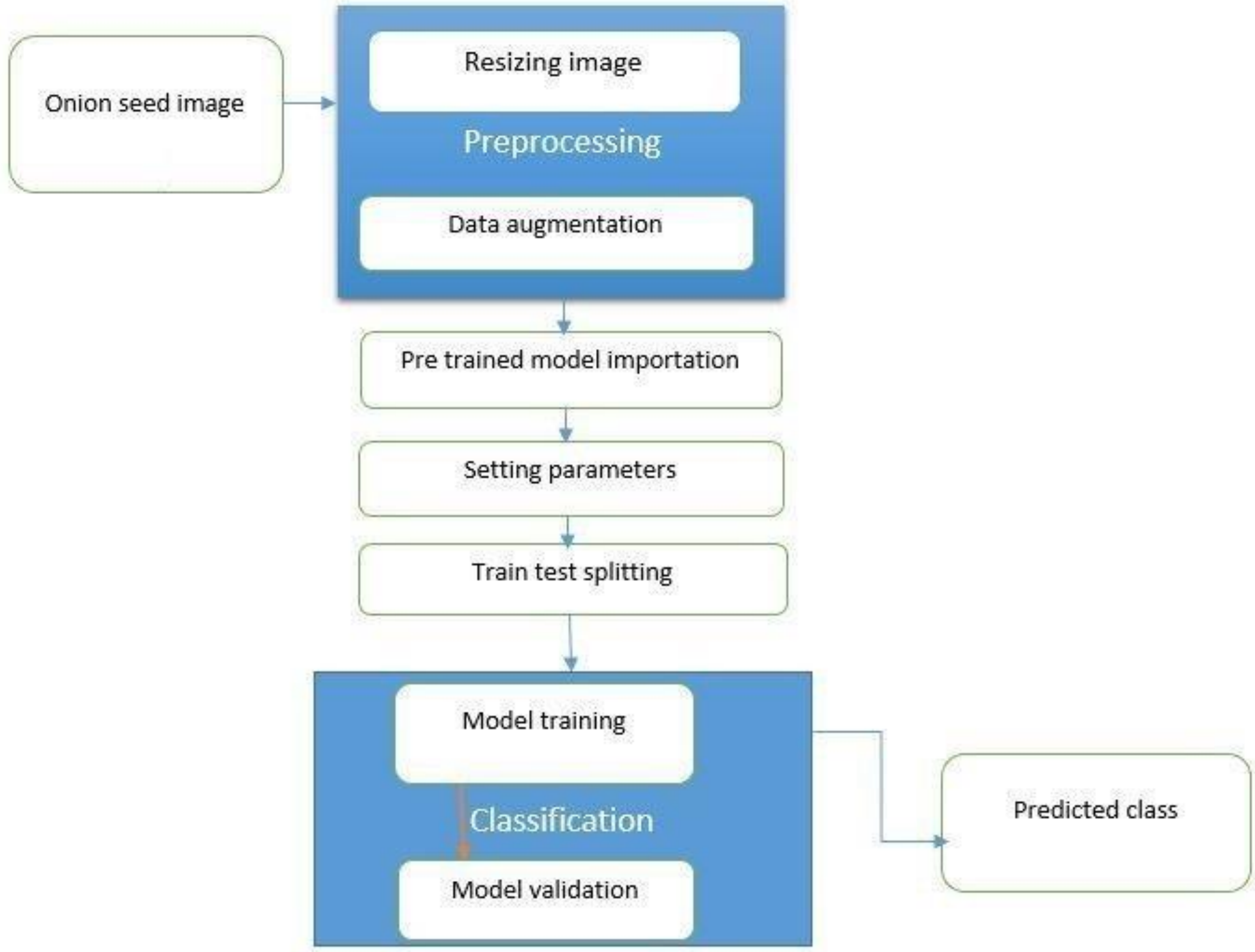

Figure 1. Proposed architecture.

Table 1. Parameters we use for the augmentation.

Image_data_generator_augmentation techniques	Values
Rotation range	40
Width shift range	0.2
Height shift range	0.2
rescale	1./255
Shear range	0.2
Zoom range	0.2
Horizontal flip	True
Brightness range	[0.2, 1.2]
Fill mode	'nearest'

3.1.4. Validation and Training of the Model

Training Parameter of the Network: A number of trials and errors were conducted in order to determine the set of training parameters that would be utilised to train the model. The Adam optimizer was employed. Additionally, 15 epochs were selected. The network's higher accuracy led to the selection of a learning rate of 0.001. The classification task was handled by SoftMax, while the activation function was handled by ReLU. The cross-entropy loss function is employed since the networks' last layer is a classification layer. The table below lists the training parameters that were used to train the model.

Table 2. Parameters we use for the work.

Parameter	Unit
Epochs	10
Batch size	256
Width, height, depth	224, 224, 3
Activation function	Softmax
Loss function	Categorical cross entropy
Optimization algorithm	Adam optimizer

Parameter	Unit
Learning rate, patience, verbose	Learning rate schedule, 25 patience, 1
Early stop, random state	5, 0

3.2. Experiment

3.2.1. Data Set

There is no predefined dataset in red onion seed for this type of research; as a result, we have developed our own dataset to evaluate the performance of the suggested system. A Sony camera (DSC-W800) is used for image acquisition in order to accomplish this. The camera was positioned 15 cm away from the sample, and all of the photos were in 24-bit color JPEG format. The images were captured with a 5152 * 3864-pixel resolution. We contrasted the colors red and blue when choosing the background color. We noticed that the blue color produced greater quality and contrasted well with the foreground elements. To train and validate the suggested model, 470 red onion seed kernels are created. The agricultural fields in the south Gondar libo kemkem and fogera woreda are the source of all the data. Agricultural specialists in certain woredas separate and categories the contents of the 470 red onion seed samples into the appropriate three classifications. As a result, there are now three categories of images: foreign, immature, and healthy.

Table 3. *Number of datasets used in each class.*

No	Class	Number of datasets
1	Healthy	150
2	Immature shriveled grain	160
3	Foreign matter	160
Total		470

The initial data has been expanded to 4000 Rotation range, width and height shift range, rescale, shear range, horizontal flip brightness range, and fill mode for class balance are some of the augmentation strategies we use to remove part of the model's 3942 images. Following data augmentation, we do train-test splitting, as seen below.

```
[[ ] print("Number of training data :", len(X_train))
     print("Number of testing data :", len(X_test))
     print("Humber of validation data :", len(x_validate))
     Number of training data : 2759
     Number of testing data : 591
     Number of validation data : 592
```

3.2.2. Implementation

The prototype was developed using the Python programming language. Google Colab is where the implementation is carried out. We selected Google Collaboratory because it is particularly well-suited for data analysis and machine learning, and it enables the writing and execution of any Python code within the browser. We also chose Colab because it is a hosted Jupyter notebook service that offers free access to computing resources, including GPUs, and doesn't require any setup. Due to the fact that Google Colab has GPUs available, we utilised them as the hardware accelerator or device. The table below lists the primary library packages utilised during implementation.

Table 4. *Libraries used in during implementation.*

Library	Version	Description
Keras	2.8.0	Deep learning model development and evaluation that is open source
Opencv	4.1.2.30	An open-source library for image processing, machine learning, and computer vision
Scikit_learn	1.0.2	Free software for data analysis in machine learning
Scikit_image	0.18.3	Open source designed for image processing

4. Result and Discussion

To find the best-performing classifier based on the classification accuracy criterion, the performance of the suggested red onion seed quality classification is evaluated using the VGG16, VGG19, GoogleNet, and ResNet50 classifiers. We present their findings in the following subsections.

4.1. VGG19 Model

VGG19 has demonstrated a 99.155% validation accuracy, a 99.49% test accuracy, and a 99.38% training accuracy. The training lasted approximately 152 minutes. The validation loss decreased from 0.5909 to 0.1148 at the conclusion of the last iteration. The validation accuracy line and the training accuracy line are nearly in sync, and the validation loss line is also in sync with the training loss, as can be seen in the two plots above. Below is the confusion matrix.

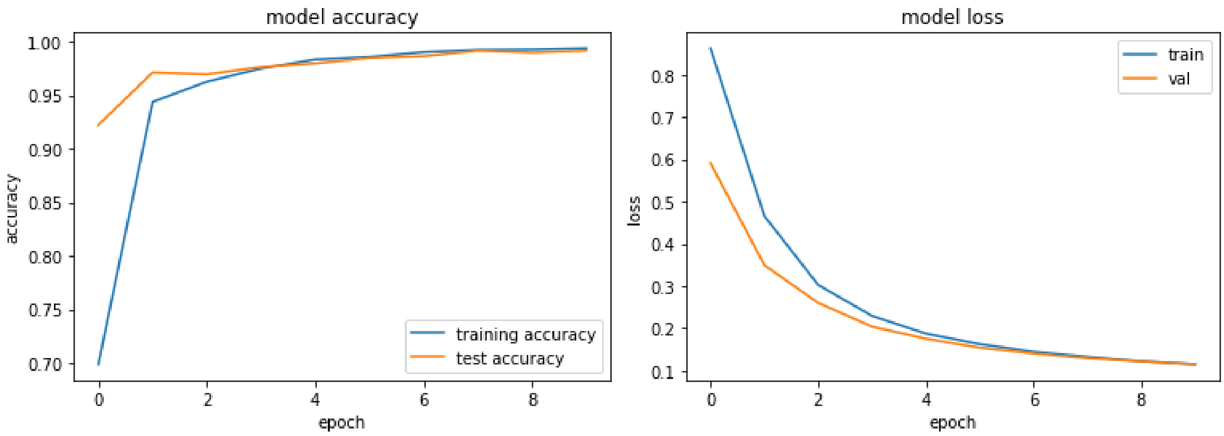

Figure 2. VGG validation accuracy and loss results.

```
              precision    recall  f1-score   support

           0       1.00      0.99      0.99       202
           1       0.99      1.00      1.00       191
           2       0.99      0.99      0.99       198

    accuracy                           0.99       591
   macro avg       0.99      1.00      0.99       591
weighted avg       0.99      0.99      0.99       591

[[200   1   1]
 [  0 191   0]
 [  1   0 197]]
```

Figure 3. VGG19 confusion matrix.

According to the confusion matrix above, there are 200 correctly predicted instances in class 0 (foreign matter) and 1 incorrectly predicted instance in each of classes 1 and 2. In class 1 (healthy), there are 191 correctly predicted instances and no incorrectly predicted instances, and in class 2 (immature), there are 197 correctly predicted instances and 1 incorrectly predicted instance in each of the classes.

4.2. VGG16 Model

VGG16 has demonstrated a 100% test accuracy, 99.93% training accuracy, and 99.831% validation accuracy. Training and validation took around 138 minutes, and by the end of the last iteration, the loss had decreased from 0.2068 to 0.0103148. Below is the confusion matrix.

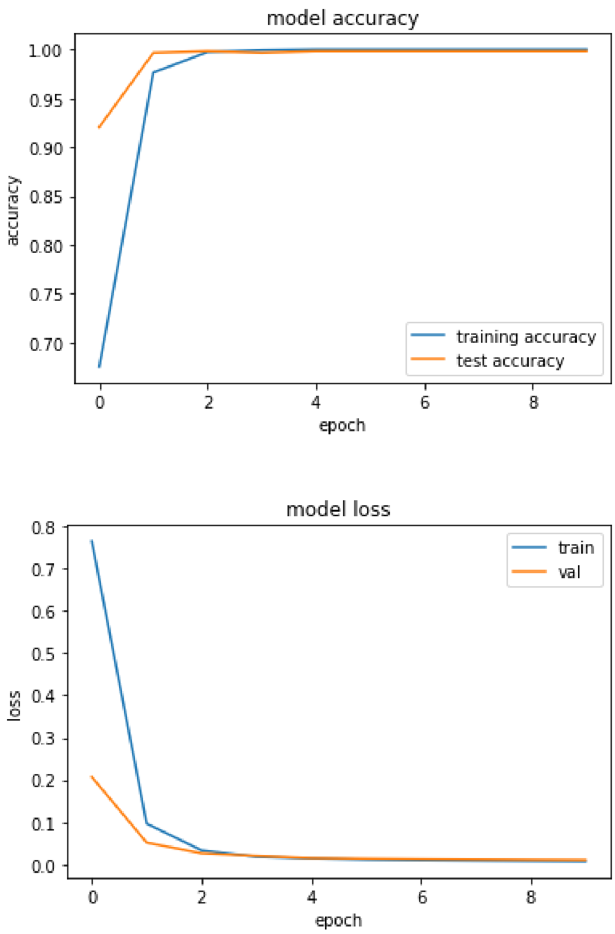

Figure 4. VGG16 validation accuracy and loss results.

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       202
           1       1.00      1.00      1.00       191
           2       1.00      1.00      1.00       198

    accuracy                           1.00       591
   macro avg       1.00      1.00      1.00       591
weighted avg       1.00      1.00      1.00       591

[[202   0   0]
 [  0 191   0]
 [  0   0 198]]
```

Figure 5. VGG16 confusion matrix.

Every instance in every class, along with its class, is accurately anticipated. Class 0 (foreign matter) has 202 correctly predicted cases with no incorrectly predicted cases, class 1 (healthy) has 191 correctly predicted cases with no incorrectly predicted cases, and class 2 (immature) has 198 correctly predicted cases with no incorrectly predicted cases.

4.3. GoogleNet Model

Validation accuracy of 99.83%, test accuracy of 100%, and training accuracy of 99.96% have all been reported by Inception V3. Training took around 142 minutes, and by the end of the last iteration, the validation loss had decreased from 0.0325 to 0.0165. The matrix of misunderstanding is displayed below in.

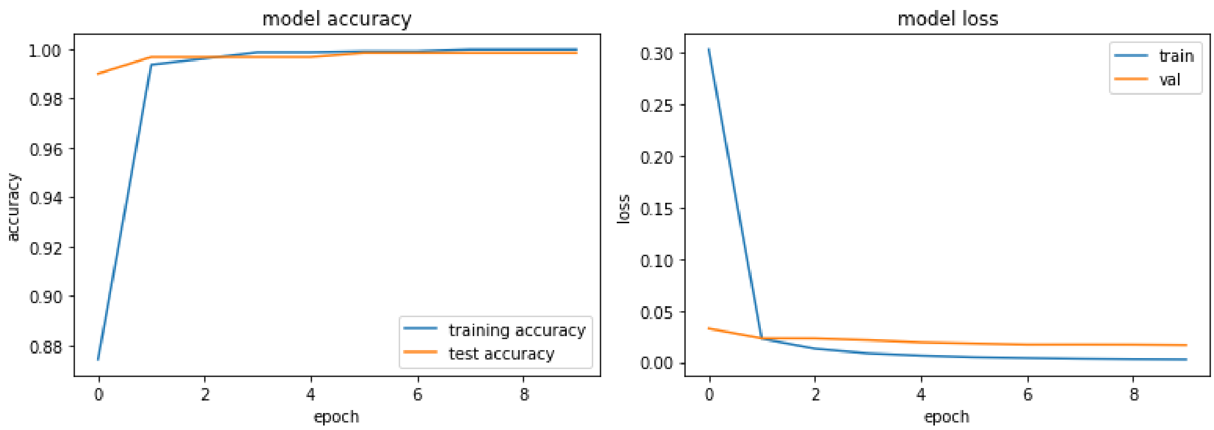

Figure 6. InceptionV3 validation accuracy and loss results.

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       202
           1       1.00      1.00      1.00       191
           2       1.00      1.00      1.00       198

    accuracy                           1.00       591
   macro avg       1.00      1.00      1.00       591
weighted avg       1.00      1.00      1.00       591

[[202   0   0]
 [  0 191   0]
 [  0   0 198]]
```

Figure 7. InceptionV3 confusion matrix.

Every instance in every class, along with its class, is accurately anticipated. Class 0 (foreign matter) has 202 correctly predicted cases with no incorrectly predicted cases, class 1 (healthy) has 191 correctly predicted cases with no incorrectly predicted cases, and class 2 (immature) has 198 correctly predicted cases with no incorrectly predicted cases.

4.4. ResNet50 Model

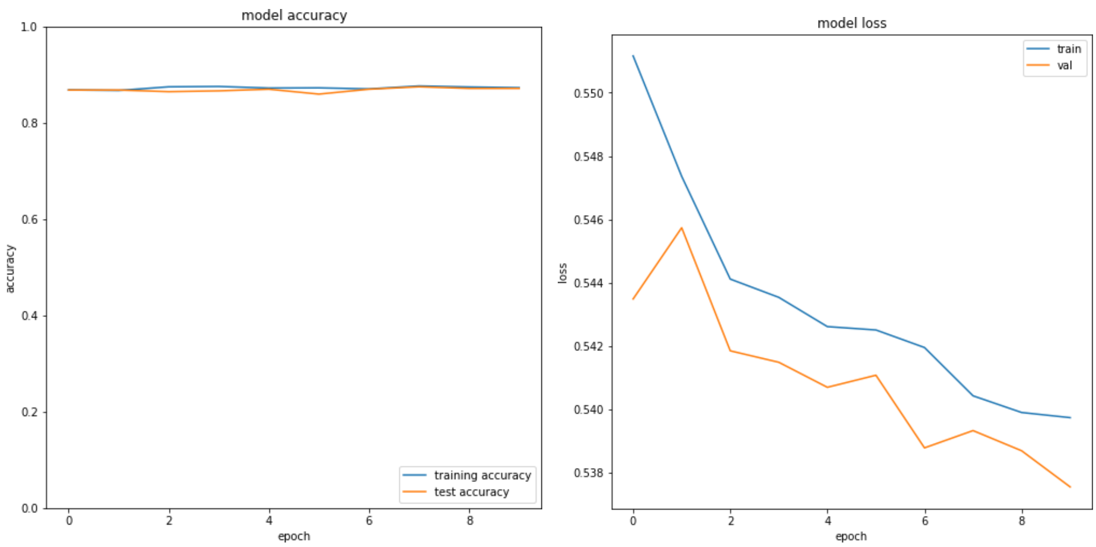

Figure 8. *ResNet50 validation accuracy and loss results.*

ResNet50 has demonstrated an accuracy of 87.11% for training, 86.29% for testing, and 87.16% for validation. Training took around 180 minutes, and by the end of the last iteration, the validation loss had decreased from 0.5435 to 0.5376. Below is the confusion matrix.

```
              precision    recall  f1-score   support

           0       0.86      0.84      0.85       202
           1       0.99      0.88      0.93       191
           2       0.77      0.86      0.81       198

    accuracy                           0.86       591
   macro avg       0.87      0.86      0.87       591
weighted avg       0.87      0.86      0.87       591

[[170   1  31]
 [  2 169  20]
 [ 26   1 171]]
```

Figure 9. *ResNet50 confusion matrix.*

The confusion matrix shows that 170 instances are correctly predicted as class 0 or foreign matter while 1 and 31 are incorrectly predicted in first and second class, respectively; 169 instances are correctly predicted as class 1 or healthy while 2 and 20 are incorrectly predicted in first and second class, respectively; and 171 instances are correctly predicted as class 2 or immature while 26 and 1 are incorrectly predicted in first class and first class, respectively.

In this study, we explore image processing approaches and supervised learning algorithms to propose quality separation for red onion seeds that functions similarly to traditional quality separation procedures. The suggested model was evaluated using verified sample data that was gathered by woreda experts from the agricultural fields of Libo Kemkem and Fogera. We evaluated the created model's performance using ac-curacy measurement. Pre-trained convolutional neural network models VGG, GoogleNet, and resnet50 were used in the trials. The accuracy rate has been used to determine the performance of the red onion quality categorization. GoogleNet, ResNet50, VGG19, and VGG16 had validation accuracy of 99.155%, 99.831, and 99.83%, respectively. We have demonstrated the effectiveness of our suggested system and the models we employed to separate the constituents of the red onion seed sample in terms of quality. This is seen by the extremely high categorization accuracy we were able to attain. VGG16 outperformed the other two classifiers in terms of validation accuracy.

Table 5. *General comparison between the results.*

Networks	Validation accuracy	Training accuracy	Testing accuracy	Average accuracy	Training time (min)
VGG19	99.155%	99.38%	99.49	99%	152
VGG16	99.831	99.93%	100%	100%	138

Networks	Validation accuracy	Training accuracy	Testing accuracy	Average accuracy	Training time (min)
GoogleNet (Inception_V3)	99.83%,	99.96%	100%	100%	142
ResNet50	87.16%	87.31%	86.29%	86%	180

5. Conclusion

An essential vegetable that is grown all over the world and utilized to flavor cuisine is the onion. Red onion seed is graded before being exported. Sorting or classifying red onion seed quality is crucial to the trading process. The main factors affecting the quality of red onion seeds should be evaluated and controlled in order to boost revenue. In this thesis, we investigated supervised learning algorithms and image processing techniques for autonomously classifying red onion seed quality using a computer system. From the definition of a picture to the most recent state-of-the-art discoveries, techniques, and structures, a comprehensive examination of image processing in general was carried out. Along with supervised learning techniques, basic processes in digital image processing are covered and briefly examined. The suggested system requires a dataset from the start. With the help of those woreda agricultural specialists, we gather the dataset from north Gondar in Libo Kemkem and Fogera woreda. In order to better and more precisely find each red onion seed kernel in the image, image preparation, including scaling and data augmentation, was completed before feature extraction from the acquired images. The primary goal of this research project is to use transfer learning techniques to create an automatic red onion seed quality assessment. We therefore suggested a system architecture that incorporates preprocessing elements such as scaling, data augmentation, and identification. We construct a labelled training dataset consisting of healthy, immature shrivel led grain, and foreign materials in order to develop a red onion seed quality classification classifier. The domain specialist from the southern Gondar libo kemkem woreda completes the labelling assignment. After all, throughout the training phase, the quality inspection model was trained on it. Training and testing datasets are separated from the image dataset. 70% of the entire dataset is used for training, while the remaining 30% is used for validation and testing using pre-trained CNN models, such as VGG (19, 16), GoogleNet, and ResNet50. For the evaluation, the accuracy evaluation technique was applied.

Abbreviations

VGG	Visual Geometry Group
GoogleNet	Google Neural Network
ResNet	Residual Neural Network
RCBD	Randomized Complete Block Design
ReLU	Rectified Linear Unit
DSC-W800	Camera Name
JPEG	Joint Photographic Experts Group
GPUs	Graphics Processing Unit
CNN	Convolutional Neural Networks
RGB	Red Green Blue

Author Contributions

Tarekegn Walle Yirdaw: Data curation, Funding acquisition, Resources, Validation

Ermias Melku Tadesse: Conceptualization, Investigation, Project administration, Writing – original draft

Endalkachew Hiwote: Formal Analysis, Supervision

Abebaw Mebrate: Methodology, Software

Ambaw Mulatu: Visualization, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] C. W. Gathambiri, W. O. Owino, S. Imathiu, and J. N. Mbaka, "Postharvest Losses Of Bulb Onion (Allium Cepa L.) In Selected Sub-Counties Of Kenya," *African J. Food, Agric. Nutr. Dev.*, vol. 21, no. 2, pp. 17529–17544, 2021, https://doi.org/10.18697/ajfand.97.20145

[2] M. A. Zaki, S. Narejo, M. Ahsan, S. Zai, M. R. Anjum, and N. Din, "Image-based Onion Disease (Purple Blotch) Detection using Deep Convolutional Neural Network," vol. 12, no. 5, pp. 448–458, 2021.

[3] T. Project, S. Horticulture, F. Empowerment, and M. Agriculture, "የቀይ ሽንኩርት አመራረት," ["Onion Production,"]2019.

[4] N. A. P. Lestari, R. Dijaya, and N. L. Azizah, "Identification Growth Quality of Red Onion during Planting Period using Support Vector Machine," *J. Phys. Conf. Ser.*, vol. 1764, no. 1, 2021, https://doi.org/10.1088/1742-6596/1764/1/012060

[5] M. H. Ahmed, "College of Natural Sciences Automatic Soybean Quality Grading Using Image Processing and Supervised Learning Algorithms," no. October, 2021.

[6] S. Afaq *et al.*, "Smart Agricultural Technology Automatic and fast classification of barley grains from images: A deep learning approach," *Smart Agric. Technol.*, vol. 2, no. December 2021, p. 100036, 2022, https://doi.org/10.1016/j.atech.2022.100036

[7] M. Habtewold, C. Dane, A. Ayza, N. Resource, M. Directorate, and A. Agricultural, "Participatory Evaluation and Demonstration of Onion Spacing in Irrigated Agriculture at Kencho Kebele in Uba Debre Tsehay Woreda, Southern Ethiopia," vol. 31, no. 2, pp. 105–114, 2021.

[8] N. Semman, G. Etana, and T. Mulualem, "Adaptability and yield performance evaluation of onion (Allium cepa L.) varieties in Jimma zone, Southwestern Ethiopia.," vol. 9, no. 4, pp. 405–409, 2019, https://doi.org/10.15580/GJAS.2019.4.090919169