INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

# Dataset Development for Automated Grade Labelling of Virginia Flue-Cured Tobacco Leaves in Tanzania: A Focus on Stalk Leaf Position

**Faith Nguleni[1,2]*, Devotha Nyambo[2], Jacob Lisuma[3], Shubi Kaijage[2]**

**1** Department of Information and Communication Technology, Tanzania Public Service College, P.o Box 329, Tabora, Tanzania
**2** School of Computation and Communication Science and Engineering, The Nelson Mandela African Institution of Science and Technology, P.o Box 447, Tengeru, Arusha Tanzania
**3** Department of Human Resource and Management, Tobacco Research Institute of Tanzania, P.o Box 431, Tumbi, Tabora, Tanzania

OPEN ACCESS

**Received:** 06-11-2024
**Accepted:** 31-01-2025
**Published:** 26-02-2025

**Citation:** Nguleni F, Nyambo D, Lisuma J, Kaijage S (2025) Dataset Development for Automated Grade Labelling of Virginia Flue-Cured Tobacco Leaves in Tanzania: A Focus on Stalk Leaf Position. Indian Journal of Science and Technology 18(7): 550-558. https://doi.org/10.17485/IJST/v18i7.3608

* **Corresponding author**.

ngulenif@nm-aist.ac.tz

**Funding:** None

**Competing Interests:** None



Published By Indian Society for Education and Environment (iSee)

**ISSN**
Print: 0974-6846
Electronic: 0974-5645

## Abstract

**Objectives:** This study aimed at developing a dataset for automated grade labeling of Virginia flue-cured tobacco leaves based on stalk leaf position by focusing on quality, colour and anomalies. **Methods**: Virginia flue-cured tobacco leaves were collected from four Tanzanian tobacco regions: Tabora municipal, Uyui, Urambo and Kaliua. Canon 5D Mark III cameras with a Canon EF 100mm F/2.8L Macro IS USM lens were used to capture tobacco leaves. The collected data concentrated on the upper leaves of the tobacco plant, also known as leaf position. In Tanzania, the tobacco plant position is a very crucial entity during grade labeling processes. **Findings**: To fulfil the study's intention, a dataset was created by collecting Virginia flue-cured tobacco leaf images. The study utilized our published dataset of Virginia flue-cured tobacco leaf images, which consisted of 49,779 high-resolution images with 22 grade labels (classes). **Novelty**: The important findings highlight the dataset's quality, making it crucial to develop automated systems for Virginia flue-cured tobacco leaf grade labeling processes based on stalk leaf position.labeling

**Keywords:** Leaf Position; Tobacco Plant; Leaves; Grade Labelling; Virginia Flue-Cured Tobacco

## 1 Introduction

Tobacco leaves are products obtained from tobacco plants; tobacco plants are grown for their leaves. The leaves are harvested from tobacco plants after they have matured enough and have been cured ready to be sold soon after being allocated with grade labels[(1)]. The process of tobacco grade labeling plays a crucial role in the sale of cured tobacco leaves, setting their price before the sale[(2)]. However, most tobacco plant cultivating countries, including Indonesia, Malawi, Kenya and Tanzania, manually grade label cured tobacco leaves on human senses[(3,4)]. Manual grade labeling of cured tobacco leaves has many drawbacks, including poor efficiency, high costs, strong

subjectivity and significant inaccuracies [5,6]. Therefore, the necessity of an automated approach cannot be escaped, as it reduces or eradicates the drawbacks; hence it increases efficiency and guarantees the unbiased grade labeling of cured tobacco leaves [7].

The existing manual methods for grade labeling encounter numerous challenges, including low accuracy in tobacco classes. Therefore, it is crucial to continue researching more intelligent and efficient methods for grade labeling of flue-cured tobacco leaves [8]. Tobacco farmers normally complain when selling their tobacco leaves as the grade labeling is done purely manually, as discussed in the study by M. Clark et al., 2020 [9]. The authors discouraged the manual method of tobacco leaf grade labeling as they struggled to quote some farmers who were saying "There are almost 10 grades of tobacco, and farmers often don't get the grade (they're) expecting." Therefore, to improve the grade labeling of their tobacco leaves, farmers have to pay a bribe to the leaf graders. As a farmer in Busia stated, "You have to bribe once you get inside there; this is Kenya, you give out money to get money", hence the only solution is an automated grade labeling system.

In Tanzania, as presented by the Tanzania Tobacco Board (TTB), tobacco plant position matters a lot, as they are among the very important entities to consider when doing grade labeling of tobacco leaves. The main tobacco plant positions used are four, the first position known as "Lugs and Primings", refers to the lowest and oldest leaves on the plant and also contains a lot of nicotine. The second position of the tobacco plant position is known as "cutters", referring to the upper leaves that follow the bottom leaves (lugs and primings). These leaves are middle-sized and contain less nicotine than lugs and primings. The third position of the tobacco plant, known as the "Thin Leaf", houses the middle tobacco leaves. The highest quality and most flavourful leaves on the stalk occupy the fourth position of the tobacco plant known as the "Leaf". The quality of these leaves is typically higher at higher leaf positions as a result of having more time to mature and develop their flavour. The leaves on this position of the plant have a stronger flavour and good grade labels [10,11].

In Tanzania, each tobacco leaf is sorted based on its similarities and differences and is tightened together in a bunch of a human hand size. The bunch must consist of similar tobacco leaves based on the leaves' position, colour, quality, length and the special factors (maturity spots, fully matured or fully orange coloured) of the leaves. The process of sorting tobacco leaves is called grading which is done by a farmer before taking his or her tobacco leaves to the marketplace so that they can be assigned a grade label before being sold. The grade labels are manually allocated to the tobacco bunch by tobacco leaf experts from the Tanzania Tobacco Board (TTB). The tobacco leaf experts need to have a minimum of three years of skills and experience in allocating tobacco grade labels before starting to allocate the grade labels on the tobacco leaves. The tobacco leaf experts, allocate the grade labels on the bunch of tobacco leaves after physically looking at each leaf on the bunch, matching their similarities to each other based on leaf position, colour, quality, length and the special factor of the leaf [10].

Figure 1 displays the grade labelled tobacco bunch samples prepared for sale in Tanzania. The sample grade labels are labeled starting with a letter that represents a plant position, followed by a number that represents a grade label quality, followed by a letter that represents a color of a leaf and finally, it follows the letter F which is used if the particular grade label has a special factor. M1L= M (Thin Leaf) is a plant position, 1 is the choice quality of the leaves, and L (lemon) is the colour of the leaves. C3O= C (cutters) is a plant position, 3 is the good quality of the leaves, and O (orange) is the colour of the leaves. L1OF= L (Leaf) is a plant position, 1 is the choice quality of the leaves, O (orange) is the colour of the leaves, and F is the factored tobacco leaves.


(a) (b) (c)

**Fig 1. Displays the grade labelled tobacco bunch samples prepared for sale in Tanzania (a) M1L (b) C3O (c) L1OF**

Datasets are crucial when it comes to machine learning models, there are thousands of data repositories on different websites that provide access to millions of datasets[12]. Many scientists and researchers face challenges of missing potential data in the datasets. Therefore, before using the datasets it is very important to identify data within the dataset to know if it meets the requirements of the task at hand. This process helps to get accurate and better results from machine learning algorithms[13]. The cured tobacco leaves dataset used[14], comprised of leaves colour and three leaves positions named upper position, middle position, and lower position. The dataset helped the authors to grade label tobacco leaves by considering the colour of tobacco leaves within three leaf positions only, the dataset could not be used on other potential features like quality and anomalies of tobacco leaves which could be used to develop more effective grade labeling results. Additionally, the dataset created by the authors during the study, cannot be accessed publicly as it can be only accessed from the corresponding authors on reasonable request. Limited access to the datasets may hinder the development of certain systems due to the lack of available data[14].

In the study of M. Lu et al., 2023 the automated system for cured tobacco leaves grade labeling was created using tobacco leaves images based on the growing position (upper part, middle part, and lower part) by focusing on maturity, oil content, structure, mutilation and identity of the tobacco leaves as the dataset used in the study was able to be used in the mentioned tasks. The dataset could not consist of other identified entities which could be used to analyse tobacco leaves such as anomalies and colour as they are very important when grade labeling the cured tobacco leaves[15]. The study of Wu et al., 2022 used the dataset of cured tobacco leaves images for grade labeling of cured tobacco leaves by analysing appearance and quality such as colour, maturity, body, oil content, colour intensity, and waste of tobacco leaves as the dataset could support that kind of analysis. The dataset could not be used to analyse the grouping of cured tobacco leaves based on their stalk leaves positions and anomalies which are crucial when grade labeling cured tobacco leaves[16].

The major role of tobacco leaf grade labeling is to maximize the utilization of tobacco resources and enhance its overall advantages. Nevertheless, it has been pointed out that the existing technology used for grade labeling cured tobacco leaves encounters obstacles. To tackle these concerns, the literature has put a forth tobacco leaves grade labeling approach that relies on the Whale Optimisation Algorithm and a multi-heterogeneous classifier stacking ensemble. However, the literature still did not explain the specific dataset used in the study as it is crucial to help in knowing whether the dataset will fit the requirements of a particular algorithm[6]. The dataset for grade labeling cured tobacco leaves focused on tobacco leaf position, green-variegated (GV), and grade label of intact tobacco leaves, the dataset could not be used to assess the quality, anomalies, and colour of tobacco leaves as they were not among the targeted data within a dataset[17]. Nevertheless, the study of Thimmegowda and C. Jayaramaiah, 2023 used the tobacco leaves dataset with a focus on a machine learning (ML) classifier that utilizes the colour, texture, and shape of tobacco leaves, as the study could not use the dataset to group cured tobacco leaves based on their leaf positions and anomalies of tobacco leaves as the dataset could not consist these entities in its analysis[18].

Despite the availability of many datasets that use deep learning techniques, the existing datasets used in developing tobacco leaf grade labeling systems are limited in their scope, often focusing on a limited set of features such as colour, quality, and position of tobacco leaf. For example, the studies of[17],[18] and others lack comprehensive data on important aspects of tobacco leaves such as anomalies and quality of a leaf. This limitation limits the effectiveness of machine learning algorithms in developing robust grade labeling systems for cured tobacco leaves. Most of the datasets did not consider the grouping of cured tobacco leaves based on stalk leaf position specifically by focusing on quality, colour, and anomalies of tobacco leaves, as are considerably viable in increasing the accuracy of cured tobacco leaves grade labeling[19]. Additionally, the availability of some of the datasets is often restricted, further hindering research and development efforts. The current research gap lies in the necessity of comprehensive and publicity accessible datasets that includes a broader range of tobacco leaf features, for developing a system for grade labeling cured tobacco leaves based on stalk leaf position with a focus on quality, colour, and anomalies of tobacco leaves, thus proposing a comprehensive solution of this study. The following methodology section, explains the processes and procedures used in collecting data required for this study and the availability of the collected data.

## 2 Methodology

### 2.1 Methods used to identify data requirements

Before heading to the field for data collection, the research team conducted a thorough investigation. This included reviewing various literature, conducting field surveys, and collaborating with the Tanzania Tobacco Board (TTB), Tobacco Research Institute of Tanzania (TORITA), and tobacco farmers.

## 2.2 Area of the study

Imagery data of Virginia flue-cured tobacco leaves were collected from four tobacco regions in Tanzania namely Tabora municipal, Uyui, Urambo, and Kaliua, as these tobacco regions are responsible for contributing more than 50% of Tanzania's Virginia flue-cured tobacco production[20]. Each tobacco region used in this study involved five farmers and to every farmer's farm, the required samples of Virginia flue-cured tobacco leaves were taken and included in the dataset.

## 2.3 Scope of the study

This study concentrated on one tobacco plant position namely the Leaf Position which is the upper part of the tobacco plant. The reason for focusing on this position is that the leaves in this position have ample time to grow, receiving abundant nutrients from the fertilizer and adequate light. Additionally, due to their superior quality and good grade labels, these leaves fetch higher prices than those from other tobacco plant positions[11]. In Tanzania, the grade labelling of cured tobacco leaves takes tobacco plant positions high consideration. Tobacco leaves are piled from tobacco plants and cured in the barns based on their plant positions to preserve the leaves' quality, flavour, and aroma as they are waiting to be allocated with grade labels and sold when completely cured[21].

## 2.4 Data collection procedures

To ensure that quality and relevant data are obtained, the research team randomly picked 20 tobacco farmers from our tobacco regions of scope. Whereby five farmers were picked from four tobacco regions, therefore five different locations within one region were used to collect the required data. The required flue-cured tobacco leaves were to be captured from their tobacco stores during data collection. So, the tobacco farmers were informed and alerted about the coming activity of data collection and how the process would be conducted. Therefore, they were told the specific dates to be visited and were requested to give full support during the whole process of data collection.

## 2.5 Data collection duration

Imagery data of Virginia flue-cured tobacco leaves were collected for two months from February 2024 to April 2024.

## 2.6 Tools used during data collection

Cured tobacco leaf images were captured with high-resolution cameras called Canon 5D Mark III with a Canon EF 100mm F/2.8L Macro IS USM lens. The images were captured under white tents, which helped to balance the nature of the climate hence helping us to get quality cured tobacco leaf images. The firewire cables were connected between the cameras and the computers, which helped cured tobacco leaf images to display directly on the computers as soon as they were captured by a camera. Consistent flat surface backgrounds covered by earthy beige sacks were used during data collection. We used external disks for storing images and computers (laptops) were used for image display and image storage. Figure 2: Illustration of the key activities conducted during the data collection process in the field. The images depict (1) capturing images of tobacco leaves, (2) arranging and organizing leaves for proper image acquisition, (3) the setup of field equipment such as cameras, tents, and generator (4) team members working collaboratively to manage data collection logistics. These activities highlight the thorough and systematic approach taken to ensure the quality of the collected dataset.

## 2.7 Leaf experts and supporting technical staff included in data collection

The Tanzania Tobacco Board (TTB) provided the research team with five tobacco leaf experts, each with over five years of experience, to assist in data collection. Additionally, there were three cameramen and two assistant people who could help in unfolding the tobacco leaves before the leaf could be captured by a camera, as some of the leaves were very curled to the extent that the tobacco leaf could not be seen properly. Figure 3: Illustration of the image capturing process during data collection in the field. A = assistant people unfolding tobacco leaves, B = the use of a white tent and flat surface during image capturing, C = Tobacco leaf expert confirming if the image is displayed on the computer, and D = Tobacco leaf expert confirming if the image displayed on the computer is the correct grade label needed.


**Fig 2. Illustration of the key activities conducted during the data collection process in the field**


**Fig 3. Illustration of the image capturing process during data collection in the field**

## 2.8 Grade labeling approach used during data collection

The five tobacco grade leaf experts selected Virginia flue-cured tobacco leaves for camera capture. For every flue-cured tobacco leaf that was graded, three experts were involved with the selection and grade labeling process, and they had to reach an agreement on the grade label to be allocated to the particular flue-cured tobacco leaves before being captured by a camera. The agreement stipulated that when two or all three experts reached a consensus on a specific grade label, a camera would capture the leaf and include it in the required data. However, if two experts disagreed on the grade label, the camera would not consider the leaf as part of the collected data. The remaining two experts utilized the computers to confirm that the computer displayed image of the flue-cured tobacco leaf bore the correct grade label for camera capture. If the captured leaf does not fit the current grade label to be captured in the process, the leaf could be taken and rechecked by all five tobacco leaf experts, and a conclusion could be made about whether the leaf should be included or not among the required data. The researcher team divided tobacco leaves experts to eradicate any unexpected biases and ensure the grade labels were included in the dataset. Through this process, the researchers acquired comprehensive tobacco leaf data with appropriate grade labels based on different characteristics such as leaf quality, leaf colour, and leaf anomalies.

## 2.9 Dataset Availability

The dataset is available for public use and may be obtained through this URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TTPLFT. The goal of the dataset is to help in the creation of automated tobacco leaf grade labelling systems based on stalk leaf position.

Figure 4 shows some of the samples of Virginia flue-cured tobacco leaf images and their grade labels as collected in the field. L1O = L (Leaf) is a plant position, 1 is a choice quality, O (Orange) is a leaf colour, L4R = L (Leaf) is a plant position, 4 is a fair quality, R (Mahogany) is a leaf colour and L3L = L (Leaf) is a plant position, 3 is a good quality, L (Lemon) is a leaf colour.

(a) (b) (c)

**Fig 4. Flue-cured tobacco image samples from the field (a) L1O (b) L4R (c) L3L**

The study[22] highlighted the characteristics of some of the grade labels in their study, such as leaf quality, leaf colour, and unique features found in the grade labels. This study applied the same approach as it used Table 1 to explain some of the selected grade label images found in Figure 4.

**Table 1. It explains more about the images used in this study which are found in Figure 4**

| Three examples of Grade Label Names (L1O, L4R and L3L) | | | | | |
|---|---|---|---|---|---|
| L | | 1 | | O | |
| L | | 4 | | R | |
| L | | 3 | | L | |
| **Section 1 (Plant position)** | | **Section 2 (Leaves Quality)** | | **Section 3 (Colour)** | |
| **Symbol** | **Meaning** | **Symbol** | **Meaning** | **Symbol** | **Meaning** |
| C | Cutters position | 1 | Choice quality | L | Lemon |
| L | Leaf position | 2 | Fine quality | O | Orange |
| M | Thin Leaf position | 3 | Good quality | R | Mahogany |
| X | Lugs and Primings position | 4 | Fair quality | | |
| | | 5 | Low quality | | |

## 3 Results and Discussion

The investigation provided the research team with insights into the process of grade labeling tobacco leaves. The requirements obtained were: first, the type of tobacco was Virginia flue-cured tobacco leaves; second, tobacco leaves should be arranged based on the stalk leaf position of the plant; in our case, we concentrated on leaf position, third, the leaf attributes for consideration were quality, colour, and anomalies. Fourth, tobacco regions in Tanzania are many, but for the study, the following regions were selected; Tabora municipal, Uyui, Urambo, and Kaliua. From the knowledge gained, the study opted to use tobacco leaf images as the potential data required for the development of an automated grade labeling system.

Therefore, after identifying the data requirement, the study proceeded to create a dataset by collecting images of Virginia flue-cured tobacco leaves from the four selected tobacco regions in Tanzania. From each tobacco region, five different tobacco farms were used to get the required flue-cured tobacco leaf images, thus a total of 20 tobacco farms were used. The research team decided to use different tobacco farms to capture a large range of variations in tobacco leaves' appearance. For this study, a purposive sampling technique was used to select the study areas, focusing on tobacco production metrics. The selected areas contribute over 50% of Virginia flue-cured tobacco production in Tanzania[20].

The main result of this study is the creation of a dataset. A dataset consists of 49,779 images of Virginia flue-cured tobacco leaves obtained from the leaf position of the tobacco plant. The imagery dataset consists of 22 grade labels (classes) of flue-cured tobacco leaves. The tobacco leaf dataset carries 22 folders each containing images of a specific grade label. The size of each image is averaging 313 KB with a total storage size of 14.9 GB. The name of each folder represents the grade label of the images within it. Images within the folders are in different numbers as it was a random image collection process, the research team took all required images of the grade labels found at a particular farm. In the farms the grade labels were not uniformly collected hence each farmer had a different quantity of tobacco leaves for particular grade labels. The dataset is intended to be used in developing a model for grade labeling tobacco leaves based on stalk leaf position. The Virginia flue-cured tobacco leaves dataset named "Tobacco leaves dataset" can be accessed through the data identification number: https://doi.org/10.7910/DVN/TTPLFT or accessed through the following Uniform Resource Locator: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/TTPLFT.

Different numbers of cured tobacco leaf images were collected for each grade label during the data collection process. The research team collected more images in certain grade labels than in others, because it was very difficult to obtain all 22 grade labels from one farmer/farm as the maximum number of grade labels that could be obtained was seven only. This made the research team take many images of the same grade labels found on a particular farm as it was very difficult to be sure if the particular grade label could be found on the next farm. Also, it was difficult to capture tobacco leaves with the camera during sunny days, especially late afternoons because the leaves were folding and breaking as were tried to unfold. Additionally, it was very difficult to obtain an adequate number of tobacco leaf images for each grade label because the dataset was collected in one crop season whereby it is difficult to obtain all leaves for all the grades with the same production level. The tobacco leaves experts, explained that different weather conditions favour different tobacco leaf grade production[12]. Table 2 shows the 22 grade labels and the total number of images in each grade label of tobacco leaves present in our dataset.

**Table 2. Tobacco leaves grade labels (22) with the total number of tobacco leaves images for each grade label as presented in our dataset**

| Grade labels (class name) | Total number of images in each grade label |
|---|---|
| L1L | 3570 |
| L10 | 1413 |
| L10F | 2021 |
| L2L | 4914 |
| L20 | 2707 |
| L20F | 3421 |
| L2R | 32 |
| L3L | 7093 |
| L3O | 5566 |
| L3OF | 3615 |
| L3R | 262 |
| L4L | 1594 |
| L40 | 3009 |
| L4R | 348 |
| L5L | 1761 |
| L50 | 1890 |
| L5R | 1341 |
| LG | 865 |
| LK | 605 |
| LLV | 2106 |
| LND | 187 |
| LOV | 1459 |
| **Total** | **49,779 images** |

In this study, the dataset of 49,779 Virginia flue-cured tobacco leaf images, categorized into 22 grade labels, provides a comprehensive resource for developing automated grade labeling systems. This dataset surpasses previous works, such as the

dataset used by[15], which included 22,322 images with only 6 grade labels, and the dataset by[14], with 11,849 images across 9 grade labels. Additionally, in study[8] used a dataset of 21,113 images with 20 grade labels. The larger and more diverse dataset in this study, collected in natural environments using high-resolution equipment, offers a distinct advantage in capturing the variability necessary for precise and reliable machine learning models. This extensive dataset allows for more accurate differentiation of leaf quality, colour, and anomalies, which are crucial for effective grade labeling.

The currently published dataset did not include one grade label found in the leaf position. This was because during data collection we failed to get the leaves of grade label L1R due to high rainfall in the farms which swept away the fertilizer on the tobacco plants, hence no tobacco leaf of this particular grade was harvested in the locations of our scope. Also, the scope of the study concentrated only on one tobacco plant position named leaf position and the other three positions were not included. While the dataset of this study is robust, future efforts should aim to address the inclusion of all 23 grade labels and additional tobacco plant positions in the dataset to further enhance the dataset's applicability. Hence this work lays a strong foundation for future research in creating automated systems that reduce or eliminate human bias in tobacco leaf grade labelling. Moreover, the dataset of this study can be integrated into machine learning models to develop more accurate and scalable grade labeling systems for tobacco leaves. The following conclusion section highlights the importance of using comprehensive datasets when developing automated systems.

## 4 Conclusion

This paper highlights the crucial need for comprehensive datasets that to be used in developing automated tobacco leaf grade labeling systems. The findings highlight the importance of having the flue-cured tobacco leaf images, as they will improve the accuracy of the models for grade labeling of flue-cured tobacco leaves. Additionally, it is advisable to identify the data required for developing any algorithm even before having the dataset as it will help to come up with robust algorithms tailored to specific grade labeling tasks.

The creation of the "Tobacco Leaves Dataset" provides a strong foundation for the advancement of machine learning models for automated grade labeling by researchers and tobacco industry partners. The development of robust algorithms that can increase grade labeling accuracy, lessen or eradicate human bias, and solve inefficiencies in conventional systems is made possible only by high quality datasets.

We encourage researchers, especially those in scientific societies, to address the challenges of developing useful grade labeling algorithms by using comprehensive datasets that include potential characteristics based on the current task at hand. The higher the quality of the datasets, the more effective the grade labeling systems will be. This will not only enhance the satisfaction of both tobacco farmers, and the tobacco buyers (companies) but also eliminate any potential corruptions and biases that may arise during manual grade labeling processes. Hence it will make the tobacco industry a better place for all stakeholders involved.

### Acknowledgements

The authors thank the Tobacco Research Institute of Tanzania (TORITA) for providing fuel for vehicle during data collection. They also acknowledge the Tanzania Tobacco Board (TTB) for allowing tobacco leaf specialists to take part in the data collection process and help identify the necessary tobacco leaves. Finally, thanks are due to Mr. Charles Mezza, Mr. Edwin Kambo, Mr. Innocent Ilomo Mr. Albert Chale, Mr. Deusdedis Mabula, Mr. Josephat Tesha, and Mr. Elimboto Muna for their help and technical guidance during data collection.

## References

1) Chen Y, Bin J, Zou C, Ding M. Discrimination of Fresh Tobacco Leaves with Different Maturity Levels by Near-Infrared (NIR) Spectroscopy and Deep Learning. *Journal of Analytical Methods in Chemistry*. 2021;p. 1–11. Available from: https://doi.org/10.1155/2021/9912589.
2) Li S, Pan X, Chen X, Zhu J, Wu B, Xie X, et al. Comparison of tobacco grading methods based on hyperspectral information. *Tob Sci Technol*. 2021;54(10):82. Available from: https://openurl.ebsco.com/EPDB%3Agcd%3A1%3A6181841/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A161257802&crl=c&link_origin=www.google.com.
3) Gowda MTT, Chandrika J. Analysis and Classification of Ripped Tobacco Leaves Using Machine Learning Techniques. In: and others, editor. Lecture Notes on Data Engineering and Communications Technologies. Singapore. Springer. 2020;p. 171–180. Available from: https://doi.org/10.1007/978-981-15-5258-8_18.
4) Marzan CS, Ruiz CR. Automated tobacco grading using image processing techniques and a convolutional neural network. *International Journal of Machine Learning and Computing*. 2019;9(6):807–820. Available from: https://www.ijmlc.org/vol9/877-V117.pdf.
5) Jimu P, Selvam GG. Tobacco Disease Detection and Classification for Grading System Using Convolutional Neural Network. *Int J Innov Sci Res Technol*. 2022;7(9). Available from: https://doi.org/10.5281/zenodo.7171504.
6) Hou KX. A WOA-Stacking Tobacco Leaf Grading Method Based on Multi-Heterogeneous Classi ers Ensemble. *Research Square*. 2023. Available from: https://doi.org/10.21203/rs.3.rs-3166140/v1.

7) TOBACCO ECONOMIC ECOSYSTEM IN INDONESIA A Scoping Study Report. 2021. Available from: https://globalactiontoendsmoking.org/wp-content/uploads/2021/09/Indonesia%20Scoping%20Study%20Final%20Report.pdf.
8) Xin X, Gong H, Hu R, Ding X, Pang S, Che Y. Intelligent large-scale flue-cured tobacco grading based on deep densely convolutional network. *Scientific Reports*. 2023;13. Available from: https://doi.org/10.1038/s41598-023-38334-z10.1038/s41598-023-38334-z.
9) Clark M, Magati P, Drope J, Labonte R, Lencucha R. Understanding alternatives to tobacco production in kenya: A qualitative analysis at the sub-national level. *Int J Environ Res Public Health*. 2020;17(6). Available from: https://doi.org/10.3390/ijerph17062033.
10) Tobacco Growing Areas. . Available from: https://www.tobaccoboard.go.tz/tobacco-growing-areas.
11) Zhang L, Zhang X, Ji H, Wang W, Liu J, Wang F, et al. Metabolic profiling of tobacco leaves at different growth stages or different stalk positions by gas chromatography-mass spectrometry. *Ind Crops Prod*. 2018;116:46–55. Available from: https://doi.org/10.1016/j.indcrop.2018.02.041.
12) Noy N, Brickley D, Burgess M. Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. In: 28th Web Conference (WebConf 2019), ACM. Assoc Comput Mach. 2019. Available from: https://research.google/pubs/google-dataset-search-building-a-search-engine-for-datasets-in-an-open-web-ecosystem/.
13) Jadhav A, Pramod D, Ramanathan K. Comparison of Performance of Data Imputation Methods for Numeric Dataset. *Applied Artificial Intelligence*. 2019;33(10):913–933. Available from: https://doi.org/10.1080/08839514.2019.1637138.
14) Xu M, Gao J, Zhang Z, Guo X. Multi-channel and multi-scale separable dilated convolutional neural network with attention mechanism for flue-cured tobacco classification. *Neural Comput & Applic*. 2023;35:15511–15529. Available from: https://link.springer.com/article/10.1007/s00521-023-08544-7.
15) Lu M, Wang C, Wu W, Zhu D, Zhou Q, Wang Z. Intelligent Grading of Tobacco Leaves Using an Improved Bilinear Convolutional Neural Network. *IEEE Access*. 2023;11:68153–68170. Available from: https://ieeexplore.ieee.org/document/10173480.
16) Wu T, Zhang Y, Gong Z, Lu D. Quantification of Tobacco Leaf Appearance Quality Index Based on Computer Vision. *IEEE Access*. 2022;10:120352–120368.
17) Lu M, Zhou Q, Chen T, Li J, Jiang S, Gao Q, et al. Qualitative Discrimination of Intact Tobacco Leaves Based on Near-Infrared Technology. *J Spectrosc*. 2021. Available from: https://doi.org/10.1155/2021/8807199.
18) Thimmegowda TGM, Jayaramaiah C. Cluster-based segmentation for tobacco plant detection and classification. *Bull Electr Eng Informatics*. 2023;12(1):75–85. Available from: https://doi.org/10.11591/eei.v12i1.4388.
19) Harjoko A, Prahara A. Image processing approach for grading tobacco leaf based on color and quality. *INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS*. 2019;12(1). Available from: https://intapi.sciendo.com/pdf/10.21307/ijssis-2019-010#:~:text=Image%20processing%20techniques%20such%20as,tobacco%20leaf%20with%2091.667%25%20accuracy.
20) Domitius K. Impact of tobacco curing on the environment andsocio-economic aspects in the Urambo District. 2021. Available from: https://www.semanticscholar.org/paper/Impact-of-tobacco-curing-on-the-environment-aspects-Katatwile/2ca8c9a14008abd637cb2ba484b14ae9af3c77ad.
21) Omara H, Odongo W, Kule EK. Adoption of environmentally friendly agricultural technologies among smallholder farmers: The case of rocket barn technology in flue-cured tobacco curing in Uganda. 2021;32(2):965–974. Available from: https://doi.org/10.1002/ldr.3765.
22) Nguleni F, Nyambo D, Lisuma J, Kaijage S. Dataset of Virginia Flue-cured Tobacco Leaf images based on stalk leaf position for classification tasks: A case of Tanzania. *Data in Brief*. 2024;56:1–6. Available from: https://doi.org/10.1016/j.dib.2024.110817.