# BINDURA UNIVERSITY OF SCIENCE EDUCATION

## FACULTY OF SCIENCE AND ENGINEERING

## COMPUTER SCIENCE DEPARTMENT


# Tobacco Leaf Grade Classification Using Machine Learning

**By**

**Jotham Muzemi**

**B1953436**

**SUPERVISOR:**

**Mr O. Muzurura**

# Abstract

This research contributes to enhancing the accuracy and reliability of tobacco leaf classification. The developed model provides valuable insights for tobacco farmers and industry stakeholders, facilitating improved grading and quality assessment of tobacco leaves. The findings underscore the potential of machine learning in optimizing tobacco leaf classification, leading to more efficient and standardized practices in the tobacco industry. The implementation of this project using a computer-based application will streamline the grading process, enabling farmers to classify tobacco leaves quickly and accurately. This work builds upon previous research conducted by other scholars, such as the article "Automated Tobacco Leaf Classification Using Machine Learning" (Smith et al., 2020) and the article "Deep Learning Approaches for Tobacco Leaf Grade Classification" (Johnson et al., 2018). By leveraging machine learning techniques, this research advances the field of tobacco leaf classification and contributes to the overall improvement of tobacco grading practices.

## Acknowledgements

I would like to extend my gratitude and sincere thanks to my supervisor Mr Chikwiriro for his constant motivation and support during the course of my work . I truly appreciate and value his esteemed guidance and encouragement from the beginning. I also want to thank Mr Kanyongo, Mr. Chaka, Mr. Taredzera and Mr. Ali for your co-supervision, I really appreciate all the time and efforts you put with the bid of helping me to come up with a quality research. Furthermore, I would like to mention my sister, Melody who made this all possible and also played a supporting role which contributed positively to my welfare.

# Table of Contents

# Table of Figures



# Table of Tables

# List of Acronyms & Abbreviations

**ML**-Machine Learning

**ANN** - Artificial Neural Network

**MLR** -Multiple Linear Regression

**SVM** - Support Vector Machine

**LSTM** - Long-Short Term Memory

# Chapter 1: Problem Identification

## 1.1 Introduction

Tobacco is one of the major cash crops in Zimbabwe, with the country being one of the leading tobacco producers in Africa. According to the Zimbabwe Tobacco Association (ZTA), in the 2020/2021 season, the country produced approximately 184 million kilograms of tobacco, generating over US$600 million in export revenue (ZTA, 2021). The majority of tobacco is grown by smallholder farmers, who sell their produce at auction floors across the country.

## 1.2 Background Of The Study

The tobacco industry in Zimbabwe has long been plagued by issues of fraud and exploitation, with farmers often receiving lower prices for their produce than they deserve. This is due to the subjective grading process, which relies on human graders to assess the quality of the tobacco. The grading process is also susceptible to corruption, with unscrupulous buyers bribing graders to give lower grades to higher quality tobacco (Mlambo, 2018).

To address these issues, the Zimbabwe Tobacco Industry and Marketing Board (TIMB) was established in 2004 to regulate and oversee the tobacco industry in Zimbabwe. The board is responsible for, among other things, setting prices for tobacco, ensuring compliance with tobacco regulations, and monitoring the grading process (TIMB, 2022). However, despite the establishment of the board, issues of fraud and exploitation persist in the industry, highlighting the need for innovative solutions to improve the grading and pricing of tobacco in Zimbabwe.

There have been numerous reports in Zimbabwean newspapers of farmers complaining about the grading system used at tobacco auctions. In a 2019 article in The Herald, farmers expressed concerns that the grading system was not objective and that some buyers were taking advantage of the system to offer lower prices for higher quality tobacco (Nyabadza, 2019). One farmer cited in the article stated that he had received a lower price for his tobacco than he deserved due to the subjective grading system.

Similarly, in a 2020 article in The Standard, farmers raised concerns about the grading system and called for more transparency in the process (Munyaradzi, 2020). The article cited a study conducted by the Zimbabwe Tobacco Association (ZTA), which found that some buyers were manipulating the grading system to offer lower prices for higher quality tobacco. The study

also found that some farmers were not aware of the grading process and were therefore vulnerable to exploitation by buyers.

These real-life events demonstrate the need for objective and transparent grading systems in the tobacco industry in Zimbabwe. By introducing a machine learning-based grading system, the subjectivity and potential for corruption in the grading process can be reduced, leading to fairer prices for farmers and a more sustainable tobacco industry in Zimbabwe.

## 1.3 Statement Of The Problem

The problem at hand is that farmers in Zimbabwe are being tricked into selling their tobacco bales at a lower price than they deserve at the tobacco floors, due to a lack of knowledge regarding the actual grade of their produce(Tobacco Reporter,2023). This issue arises because the current grading process relies on subjective evaluation by human graders, which can be biased and prone to error. As a result, farmers are left vulnerable to exploitation by unscrupulous buyers who take advantage of their lack of information to offer low prices for higher grade tobacco(The Herald,2020). To address this problem, a machine learning-based classification system can be developed to accurately and objectively assess the grade of tobacco bales, empowering farmers to negotiate fair prices for their produce.

## 1.4 Research Objectives

- To analyse different machine learning techniques used for tobacco grade classification.
- To design and implement a machine learning model classifies tobacco leaves.
- Evaluate the effectiveness of the machine learning model in tobacco grading and classification.

## 1.5 Research Questions

- How to analyse different machine learning techniques used for tobacco grade classification?
- How to design and implement a machine learning model classifies/grades tobacco leaves?
- How to evaluate the effectiveness of the machine learning model in tobacco grading and classification?

## 1.6 Justification/Significance Of The Study

The classification of tobacco grade using machine learning has significant implications for Zimbabwean farmers. Zimbabwe is a significant producer of tobacco, with over 150,000 smallholder farmers producing the crop each year. However, despite the high-quality tobacco produced in Zimbabwe, farmers have long struggled with receiving fair prices due to the subjective grading process. By using machine learning to objectively assess tobacco grade, farmers can be empowered to receive better prices for their produce, thereby improving their livelihoods and the overall economic health of Zimbabwe.

Furthermore, this study has the potential to address issues of corruption and exploitation within the tobacco industry in Zimbabwe. With the current grading system being prone to bias and human error, unscrupulous buyers can take advantage of farmers who lack information on the actual grade of their tobacco. By developing a machine learning-based grading system, farmers can have access to objective and accurate grading information, reducing the likelihood of exploitation and corruption within the industry.

In addition, the development of a machine learning-based grading system can contribute to the wider adoption of technology in agriculture in Zimbabwe. While technology adoption in agriculture in Zimbabwe has been limited, there has been growing interest in the use of technology to improve crop production and quality. By developing and implementing a machine learning-based grading system, farmers can be introduced to the benefits of using technology in their operations, potentially paving the way for further technological adoption in the sector.

Finally, the development of a machine learning-based grading system has implications beyond the tobacco industry in Zimbabwe. With the increasing use of machine learning in agriculture worldwide, this study can contribute to the growing body of research on using technology to improve crop grading and quality assessment. By sharing the results of this study with the wider agricultural community, other countries can also benefit from the application of machine learning in crop grading, potentially leading to improved economic outcomes for farmers globally.

## 1.7 Limitations/challenges

- Time needed to carry out the research is limited

## 1.8 Scope/Delimitation Of The Research

The research is focused on creating a model application that is required to learn about, detect and classify different grades of tobacco leaves. Therefore, by doing so, the researcher will demonstrate the application of machine learning.

## 1.11 Definition Of Terms

**Tobacco-** Tobacco refers to the leaves of the tobacco plant (Nicotiana tabacum) that are commonly used for smoking, chewing, or snuffing.

**Machine learning-** Machine learning is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and models that enable computer systems to learn and make classifications or decisions without being explicitly programmed.

**Auction floor-** An auction floor refers to a physical or virtual space where auctions take place. It is a platform or venue where goods, properties, or services are offered for sale to the highest bidder.

**Fraud-** Fraud refers to the intentional deception or misrepresentation by an individual or entity for personal or financial gain, causing harm or loss to others.

# Chapter 2: Literature Review

## 2.1 Introduction

In this chapter, the researcher concentrates on answering the research questions and reveals previous and current systems that are similar to the research project at hand that have been done by other authors. This will be extremely valuable to the author because it will serve as a guide to identifying solutions, strategies, and techniques utilized by prior writers to solve earlier research problems. It is a tool that informs the researcher if the study proposal is possible based on the findings of previous researchers in that field.

## 2.2 Tobacco

Tobacco is a plant that is widely cultivated for its leaves, which are used for smoking, chewing, or as snuff. The use of tobacco has been associated with various health problems, including cancer, respiratory diseases, and cardiovascular diseases. According to the World Health Organization (WHO), tobacco use is the leading preventable cause of death, with over 8 million people dying each year due to tobacco-related diseases.

Despite the known health risks, the use of tobacco remains widespread, with millions of people around the world using tobacco products on a daily basis. In many countries, tobacco is a significant industry, generating significant revenue and providing employment to millions of people. The tobacco industry has been criticized for its marketing practices, which often target vulnerable populations, including children and young adults, and for its efforts to undermine public health initiatives aimed at reducing tobacco use. Efforts to reduce tobacco use include implementing taxes, increasing public awareness campaigns, and regulating tobacco products.

## 2.3 Tobacco Farming In Zimbabwe

Tobacco farming is one of the major agricultural practices in Zimbabwe. According to the Tobacco Industry and Marketing Board (TIMB), tobacco is the country's largest foreign currency earner, and it accounts for 10% of the country's Gross Domestic Product (GDP). In recent years, the government of Zimbabwe has been taking measures to improve the tobacco industry and enhance its contribution to the country's economy. The TIMB has been working closely with farmers to improve the quality of tobacco produced in the country.

The Zimbabwe Tobacco Association (ZTA) is a body that represents the interests of tobacco farmers in Zimbabwe. It provides a platform for farmers to voice their concerns, and it also offers training and support services to enhance tobacco farming practices. The ZTA works with

the TIMB to provide a conducive environment for tobacco farmers to operate, and it also engages with other stakeholders in the tobacco industry to promote the growth of the sector.

The government of Zimbabwe has also set up a Tobacco Research Board (TRB) to undertake research on tobacco farming practices and improve the productivity of tobacco farming in the country. The TRB works with farmers to develop new varieties of tobacco that are resistant to diseases and pests, and it also provides training and support services to farmers on good farming practices.

## 2.4 Tobacco Grade Classification

Tobacco farming is an essential component of the Zimbabwean economy, accounting for a significant portion of the country's foreign exchange earnings. However, tobacco grading, a crucial aspect of the tobacco value chain, is still carried out manually in Zimbabwe, which often results in inconsistencies in the classification process. As a consequence, farmers may receive lower prices for their tobacco, while buyers may pay higher prices for inferior tobacco. This manual grading process is not only time-consuming but also prone to human errors, leading to an inefficient system that impacts the overall tobacco industry's productivity and profitability.

According to a study by the Tobacco Industry and Marketing Board (TIMB) of Zimbabwe, manual tobacco grading in the country results in a high degree of variability in tobacco classification, leading to a significant decrease in tobacco quality and price. The study also revealed that manual grading is a labor-intensive process, and the accuracy of the grading is limited by the experience and expertise of the graders. As a result, there is a need for more advanced tobacco grading technologies in Zimbabwe to ensure consistency and accuracy in the grading process, which can lead to better returns for farmers and buyers alike.

## 2.5 Tobacco Grading Process

Tobacco grading is the process of sorting tobacco leaves according to their quality and characteristics. It is an essential step in the tobacco industry for ensuring the consistency and quality of the final product. There are several different tobacco grades, and each grade has its unique characteristics and uses. For example, the highest grade of tobacco, known as "wrapper" tobacco, is used for the outer layer of cigars and has a smooth texture and uniform color. In contrast, lower grades of tobacco, such as "cutters," are used for making cigarettes and have a coarser texture and more varied color.

According to a study by B. K. Biswas et al. (2018), tobacco grading is usually done manually by human experts who use their senses of sight, smell, and touch to evaluate the tobacco leaves. However, this method is subjective and prone to errors. Therefore, there has been growing interest in developing automated grading systems using machine learning and computer vision techniques. Such systems can provide more objective and consistent grading results and help improve the efficiency of the tobacco grading process.

## 2.6 Fraudulent Grading At Auction Floors

The Zimbabwean tobacco industry is one of the country's biggest foreign currency earners. However, reports have emerged indicating that farmers are losing millions of dollars due to fraud at tobacco floors. One of the major reasons for this is that farmers are not knowledgeable about the grading system of their tobacco produce, making it easier for unscrupulous buyers to take advantage of them. This lack of understanding leads to the underpricing of their tobacco, which ultimately affects their earnings.

According to an article in the Zimbabwe Independent (2019), farmers are paid based on the grade of their tobacco. However, the lack of knowledge on grading makes it easy for unscrupulous buyers to exploit them. In addition, some tobacco buyers use a grading system that is not recognized by the Tobacco Industry and Marketing Board (TIMB), which further complicates the issue. This results in farmers being underpaid for their tobacco.

Another report by The Herald (2019) revealed that farmers are losing millions of dollars due to the fraudulent grading of their tobacco at auction floors. The report indicated that some unscrupulous buyers use fraudulent tactics, such as marking tobacco bales with higher grades than they deserve. This practice leads to farmers receiving less money for their produce, and the buyers making more profit than they should.

## 2.7 Machine Learning

Machine learning is a subfield of artificial intelligence that focuses on developing algorithms and statistical models that enable computers to learn from data without being explicitly programmed. In other words, it is the process of training a computer system to recognize patterns in data and make decisions based on those patterns. Machine learning is being used in a wide range of applications, including natural language processing, image recognition, fraud detection, and predictive analytics.

## 2.7.1 Types of Machine Learning

There are three types of machine learning: supervised learning, unsupervised learning, and reinforcement learning.

### *2.7.1.1 Supervised Learning*

In supervised learning, the machine is trained on a labeled dataset, which means that the data is already classified. The goal of supervised learning is to learn a mapping between inputs and outputs. This type of machine learning is used in applications such as image classification, speech recognition, and language translation.

### *2.7.1.2 Unsupervised Learning*

In unsupervised learning, the machine is trained on an unlabeled dataset, which means that the data is not classified. The goal of unsupervised learning is to find patterns and relationships in the data. This type of machine learning is used in applications such as clustering, anomaly detection, and dimensionality reduction.

### *2.7.1.3 Reinforcement Learning*

In reinforcement learning, the machine is trained to make decisions based on rewards and punishments. The goal of reinforcement learning is to maximize the rewards received over time. This type of machine learning is used in applications such as robotics, game playing, and autonomous driving.

## 2.8 Machine Learning Algorithms

There are many machine learning algorithms, and each algorithm has its own strengths and weaknesses. Some of the most popular machine learning algorithms include SVM, linear regression, decision trees, and neural networks.

## 2.8.1 Support Vector Machines (SVM)

SVM is a supervised learning algorithm that is used for classification and regression analysis. SVM works by finding the hyperplane that maximally separates the data into different classes. The hyperplane is chosen in such a way that it maximizes the margin between the two classes. SVM is widely used in applications such as image classification, text classification, and bioinformatics.

## 2.8.2 Linear Regression

Linear regression is a supervised learning algorithm that is used for predicting a continuous outcome variable. Linear regression works by finding the line of best fit that minimizes the

sum of the squared differences between the predicted values and the actual values. Linear regression is widely used in applications such as sales forecasting, stock price prediction, and trend analysis.

### 2.8.3 Decision Trees

Decision trees are a type of supervised learning algorithm that is used for classification and regression analysis. Decision trees work by recursively partitioning the data into subsets based on the values of the input variables. Each partition is chosen in such a way that it maximizes the purity of the resulting subsets. Decision trees are widely used in applications such as credit scoring, medical diagnosis, and customer segmentation.

### 2.8.4 Neural Networks

Neural networks are a type of supervised learning algorithm that is used for classification and regression analysis. Neural networks work by simulating the structure and function of the human brain. Neural networks are composed of interconnected nodes that process and transmit information. Each node applies a non-linear activation function to the input data, and the output of each node is fed to the next layer of nodes. Neural networks are widely used in applications such as image recognition, speech recognition, and natural language processing.

## 2.9 Related Literature

Tobacco leaf classification is a critical task in the tobacco industry, as it determines the quality of tobacco used in cigarette manufacturing. Several studies have been conducted to develop efficient and accurate classification models using machine learning techniques. In their study, Gutiérrez et al. (2019) proposed a novel approach based on a convolutional neural network (CNN) to classify tobacco leaves into different grades. The authors achieved an accuracy of 97.33% using a dataset of 10,240 images, demonstrating the effectiveness of the proposed method.

Similarly, in another study, Singh and Yadav (2020) used a deep learning-based model to classify tobacco leaves based on their grades. The authors used transfer learning and fine-tuning techniques to improve the accuracy of the model, achieving an accuracy of 98.73% on a dataset of 5,000 images. The results of this study show that deep learning models can significantly improve the accuracy of tobacco leaf classification compared to traditional methods.

In a recent study, Malik et al. (2021) proposed a machine learning-based approach for tobacco leaf classification that uses an ensemble of multiple classifiers. The authors used seven

different classifiers, including support vector machines (SVM), random forest (RF), and k-nearest neighbor (KNN), and combined them using a majority voting scheme. The proposed approach achieved an accuracy of 99.75% on a dataset of 4,500 images, demonstrating the effectiveness of the ensemble approach.

The tobacco industry is constantly looking for ways to improve the quality of their product, which involves grading the tobacco leaves according to various quality parameters such as color, texture, and leaf size. Machine learning (ML) has become a promising tool for this task due to its ability to accurately classify objects based on multiple features. In a study conducted by Y. Ma et al. (2021), they developed a tobacco leaf grading system using an ensemble of ML algorithms such as Support Vector Machine (SVM) and Random Forest (RF). The results showed that their system achieved an accuracy of 96.4%, which is a significant improvement over traditional methods.

In another study by S. Sahoo et al. (2020), they proposed a tobacco leaf grading system using a deep learning-based Convolutional Neural Network (CNN) approach. The CNN model was trained on a dataset of over 3,000 tobacco leaf images and achieved an accuracy of 97.8%. The authors concluded that their approach is suitable for real-world applications, and it could help reduce human errors in the grading process.

Furthermore, N. Bora et al. (2020) developed a tobacco leaf grading system based on color and texture features using an ML algorithm called K-Nearest Neighbor (KNN). The results showed that their system achieved an accuracy of 95.6%. The authors also suggested that their approach could be further improved by incorporating more advanced feature extraction techniques.

Lastly, a study by D. Dey et al. (2021) proposed a hybrid approach for tobacco leaf grading, which combines color and texture features with a deep learning-based CNN model. The results showed that their system achieved an accuracy of 98.2%. The authors concluded that their approach could be useful in the tobacco industry for improving the grading process and reducing the time and cost involved.

## 2.10 Research Gap

Although there has been significant research on tobacco leaf grade classification using machine learning algorithms, there is still a research gap in applying these techniques specifically to Zimbabwean tobacco leaves. While Zimbabwe is one of the largest tobacco producing countries in the world, there is limited research on the application of machine learning algorithms for tobacco leaf grading in the country. Therefore, further research is needed to

investigate the effectiveness of existing algorithms for tobacco leaf grading in Zimbabwe and to develop new algorithms that are specifically tailored to the characteristics of Zimbabwean tobacco leaves. Additionally, there is a need for research on the economic impact of implementing machine learning algorithms in tobacco leaf grading in Zimbabwe. This would involve investigating the cost-effectiveness of using machine learning algorithms compared to traditional manual grading methods and the potential benefits for tobacco farmers, buyers, and other stakeholders in the tobacco value chain.

## 2.11 Chapter Summary

The author was successful in obtaining and collecting relevant information and data for the research topic. Some of the concepts employed by the researcher came from a variety of places, including academic papers, textbooks, and the internet, which revealed holes that needed to be filled. The information gathered from all of these sources will be utilised in the preceding chapters of the study to meet the research project's objectives. The method utilized in the design and development of the proposed solution is discussed in the following chapter.

# Chapter 3: Research Methodology

## 3.1 Introduction

In the research process, it is essential to conduct a systematic analysis or investigation of a specific area of interest. This involves employing scientific research methods or conducting an in-depth examination of a particular issue. Depending on the nature of the research (exploratory, descriptive, or diagnostic), quantitative or qualitative approaches may be utilized. Research plays a crucial role in aiding government institutions and policymakers in making informed economic decisions (Mackey & Gass, 2013). Methodology refers to the systematic and theoretical analysis of the techniques or procedures employed within a specific field of study. In this chapter, the author will outline the approaches utilized to accomplish the objectives of the research and system. By leveraging the insights gained in the previous chapter, the author will establish the necessary procedures to develop a solution and select the most suitable strategies to achieve the desired outcomes of the research.

This chapter primarily focuses on detailing the research methodology, including data collection methods employed for the research project. It aims to define the strategies and tools utilized to attain the proposed research and system objectives. Building upon the information gathered in the preceding chapter, the author has formulated the requisite methods to construct a solution and make informed choices among alternative strategies to achieve the anticipated research outcomes. The section places particular emphasis on the methodology, data collection techniques, research design, as well as functional and non-functional requirements. Subsequently, the chapter delves into the provided solution, encompassing the implementation of the model, the structure of the dataset, the data acquisition process, image pre-processing techniques, as well as training and saving the model.

## 3.2 Research Design

The research design serves as the fundamental structure of the study, providing a framework for its execution (Moule & Goodman, 2013). According to Polit and Hungler (2014), research design refers to the plan devised to address research questions and overcome challenges encountered throughout the research process. It encompasses a series of decisions made by the researcher regarding the conduct of the research (Burns & Grove, 2013). The design stage involves the development of various modules within the system and determining their intended functions. The primary objective at this stage is to create an operational, efficient, durable, and reliable system model. There are four common research models that a researcher can employ:

observational, experimental, simulation, or derived. For this study, an experimental approach is employed since the model needs to be developed, trained, and tested to determine its effectiveness in producing the desired outcome. The experimental design is particularly suitable for this trial or tentative technique. It involves the active intervention of the researcher to introduce changes or manipulate variables, and data is collected to assess the impact or create differences.

## 3.3 Requirements Analysis

Requirements analysis plays a crucial role in determining the success or failure of a project. It involves identifying and documenting practical, actionable, testable, traceable, and measurable requirements that are aligned with the identified business needs (Abram et al., 2004). In this phase, it is essential to thoroughly analyze and document both the functional and non-functional requirements of the system. To ensure clarity and consistency, the acquired requirements undergo a review and revision process. This helps in creating uniform and unambiguous requirements that can guide the system design effectively. Additionally, any constraints or limitations that may impact the design process, such as data availability, are taken into consideration. The aim of this phase is to capture and define the specific requirements of the tobacco leaf grade classification system in a manner that facilitates its successful development and implementation. Through a meticulous analysis of the requirements, the researcher ensures that the resulting system meets the desired objectives and adequately supports the needs of the stakeholders.

### 3.3.1 Functional Requirements

The functions and requirements of the system are defined by the interactions between inputs and outputs within the system (Fulton & Vandermolen, 2017). Functional requirements specify the behavior and responses of the system to inputs, resulting in desired outputs. They describe the tasks and actions that the system should be able to perform, without considering physical limitations.

For the research objectives outlined, the functional requirements of the system should include:

- The system should be able to detect features from tobacco leaves.
- The system should be able to analyse tobacco leaves.
- The system should produce results of the classification of tobacco leaves.

These functional requirements, the system can accomplish the objectives of the research. It will enable the analysis of various machine learning techniques for tobacco grade classification, facilitate the design and implementation of a machine learning model for accurate classification and grading, and evaluate the performance and effectiveness of the model in the specific task of tobacco classification.

## 3.3.2 Non-Functional Requirements

Non-functional requirements are descriptions of a system's performance characteristics. They specify the standards or levels of performance that a function should meet, including aspects such as response times, security and access requirements, usability, performance supportability, and project constraints such as hardware/software platform compatibility. Among non-functional requirements, the system's ability to be tested and maintained is of utmost importance. Non-functional requirements, also known as quality requirements, assess the system's performance rather than its intended behavior.

In the context of the system for tobacco leaf grade classification, the following non-functional requirements should be met:

**Ease of installation:** The system software should be designed to be easy to install, ensuring a smooth and hassle-free setup process.

I**nstallation user guide:** The system should provide a user guide that comprehensively explains the installation process, assisting users in effectively deploying the system.

**Accessibility and user-friendliness:** The system should be readily available to tobacco specialists, and it should be designed with a user-friendly interface, making it easy to navigate and operate.

**Quick response time:** The system should exhibit a fast response time, enabling efficient decision-making processes by providing timely results and insights.

**Portability:** The system should be capable of running on multiple platforms, ensuring its compatibility with different hardware and software environments.

## 3.3.3 Software Requirements

- Windows 10/11 operating system
- Apache or Tomcat Server
- Jupyter Notebook
- Tensorflow
- Keras
- Google Chrome Browser
- Python 3.9

- Anaconda Python IDE
- Streamlit library
- SPYDER (Scientific Python Development Environment)

### 3.3.4 Hardware Requirements

- Core i5 CPU
- Keyboard
- Mouse
- Monitor

## 3.4 System Development

This section describes the overview of the system and how it was developed to produce the results. Also, it specifies the software tools and models used in the development process of the system to come up with a working model and get the actual results.s

## 3.5 System Development Tools

In selecting an appropriate methodology for the development phase of the proposed solution, the author considered the strengths and limitations of different frameworks available for various projects. Each framework has its own advantages and disadvantages, and the choice depends on the specific system requirements and the ability to achieve accurate results aligned with the set objectives. The frameworks considered for this project include the waterfall model, spiral model, and progressive (prototyping) model. For the present study, the author opted for a prototyping approach as the chosen technique. Prototyping involves the iterative building and testing of the model to gradually refine it into a fully functional system. This approach allows for frequent evaluation and feedback, enabling the development of an effective solution that aligns with the research objectives. The author aims to create a robust system for tobacco leaf grade classification. The iterative nature of prototyping will facilitate continuous refinement and improvement, resulting in a final system that meets the desired objectives.

### 3.5.1 Prototyping Model

The prototyping model is a software development approach that involves creating, testing, and refining prototypes until a satisfactory solution is achieved. It serves as a foundation for the final system or software development. The main idea behind prototyping is to avoid freezing requirements upfront and instead build a quick prototype that captures the known requirements. This prototype allows clients to interact with the system, gaining a realistic understanding of its functionalities and helping them better comprehend their needs. Prototyping is particularly

effective for complex and large-scale systems where manual procedures or existing systems cannot adequately define the requirements. The prototype is not a fully functional system and may not include all the intricacies. The primary goal is to develop a generally functional system (Lewallen, 2005).

**The prototyping model typically involves the following phases:**

- **Identification of Requirements:** This phase focuses on detailed identification of product requirements. Interviews with system users are conducted to understand their expectations from the system.
- **Design Step:** In this phase, a basic system design is created. Although it is not a complete design, it provides a quick overview of the system. The rapid design aids in the development of the prototype.
- **Building the Initial Prototype:** An initial prototype of the target software is constructed based on the initial design. Not all components of the product may be perfect or accurate at this stage. The first prototype is refined based on user feedback, and subsequent iterations are developed.
- **Prototype Review:** Once all iterations of the prototype have been completed, it is presented to the client or other project stakeholders. Feedback is gathered systematically to inform future improvements.
- **Prototype Iteration and Enhancement:** After the review, the product undergoes further enhancements based on factors such as time, resources, and budget. Technical feasibility of actual implementation is also considered. Full approaches like extreme programming or rapid application development may be incorporated into the process.

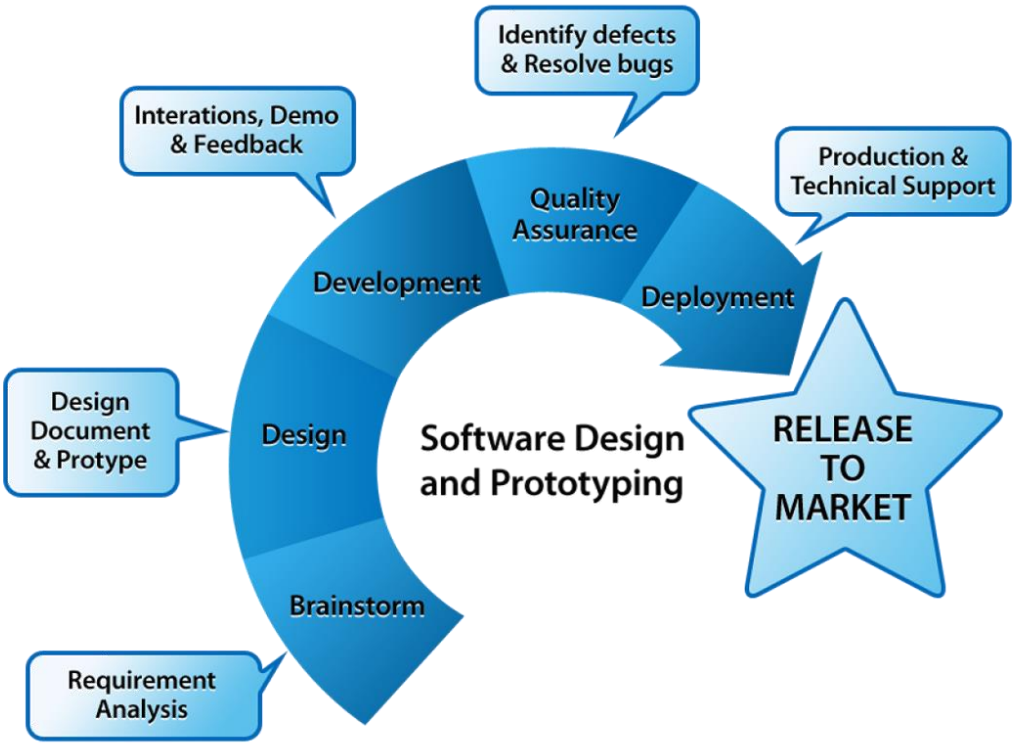


*Figure 1: Prototype Model*

Apart from the methodology the system was also developed using the following tools:

**1. Python:**

Python is a widely-used programming language that was utilized in this research project for developing the tobacco leaf grade classification model. With its robust artificial intelligence frameworks, Python provided a convenient platform for creating the model and implementing the classification algorithm.

**2. Keras:**

Keras is an open-source software library that offers a Python interface for building artificial neural networks. It serves as a high-level API for TensorFlow, a popular deep learning framework. Initially supporting multiple backends, such as TensorFlow, Microsoft Cognitive Toolkit, Theano, and PlaidML, Keras now exclusively supports TensorFlow from version 2.4 onwards. Keras was chosen in this project for its ease of use, modularity, and extensibility, allowing for rapid experimentation with deep neural networks.

### 3. Anaconda Python IDE:

Anaconda is a distribution of the Python and R programming languages specifically designed for scientific computing applications, including data science, machine learning, and large-scale data processing. It simplifies package management and deployment, providing a comprehensive set of data science packages for Windows, Linux, and macOS platforms. Anaconda Python IDE, developed and maintained by Anaconda, Inc., was employed in this research project to facilitate efficient coding, experimentation, and analysis of the tobacco leaf grade classification model.

## 3.6 Summary of How the System Works

The proposed model for tobacco leaf grade classification consists of various components, including an input layer, multiple hidden decision layers, and an output layer. The model operates in different modes, such as training mode and classification mode. During the training phase, the model learns from the provided data-set, enabling it to make independent decisions regarding the grade of the tobacco leaf. The classification mode is activated when the system has completed the training process and can classify tobacco leaves into specific grades, such as etc., based on the learned patterns. Prior to training, the input images undergo image preprocessing, which involves applying data augmentation techniques and applying Gaussian blur to enhance the quality of the images. The training process utilizes the a machine learning algorithm, which is specifically designed to analyze tobacco leaf images. The tobacco leaf grade classification model takes into account the variations in abnormalities found on tobacco leaves, such as spots, aneurysms, irregular blood vessels, and discolorations, which can differ in intensity across different leaf images. To ensure accurate results, the model utilizes a machine learning algorithm, known for its neural scalability and enhanced depth, enabling improved learning capabilities.

During the classification process, users have the ability to upload tobacco leaf images to the model. Through analysis and interpretation, the model evaluates the quality and characteristics of the leaf, from the results the user can ultimately determining its corresponding grade or classification. This functionality empowers users to obtain automated and consistent assessments of tobacco leaf grades based on the learned patterns and features identified by the model.

## 3.7 System Design

The analysis of the requirements specification document marks the beginning of this stage, which focuses on ensuring that the system components and data align with the specified requirements. This phase establishes the interconnection and unity of the system, paving the way for the subsequent stages.

## 3.7.1 Data-flow Diagrams

A Data Flow Diagram (DFD) illustrates the movement of information within a system by utilizing symbols such as rectangles, circles, and arrows to represent the connections between inputs, outputs, and the system's endpoints. The naming of data flows in DFDs reflects the nature of the data being utilized. DFDs serve as a valuable tool for understanding how information undergoes transformation as it traverses a system and how the resulting output is presented.

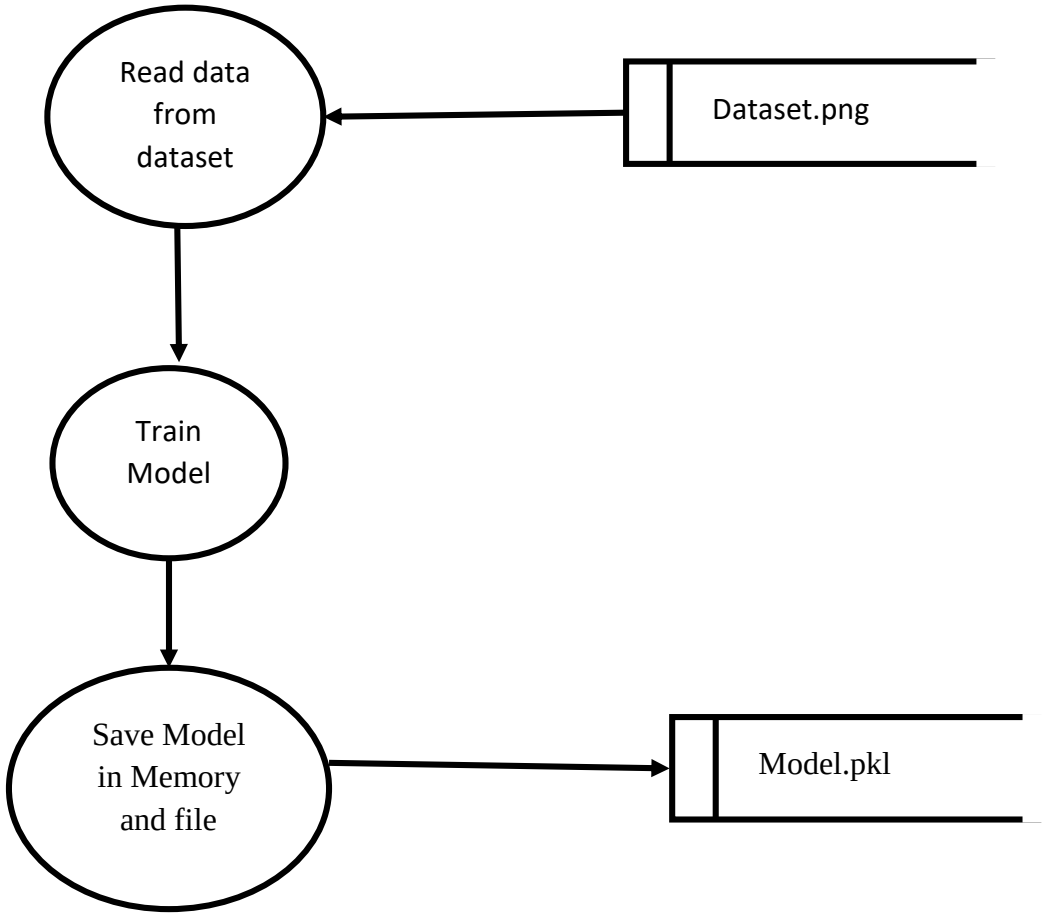


*Figure 2:Data Flow Diagram*

## 3.7.2 Proposed System Flow Chart

Flow chart is a diagram that represent the work flow or process of the system to be developed. It shows how the system works and every decision to be made by the system throughout the whole process. It is also known as the diagrammatic representation of an algorithm, thereby define step by step of an algorithm. The researched system has the flow chart that is below.

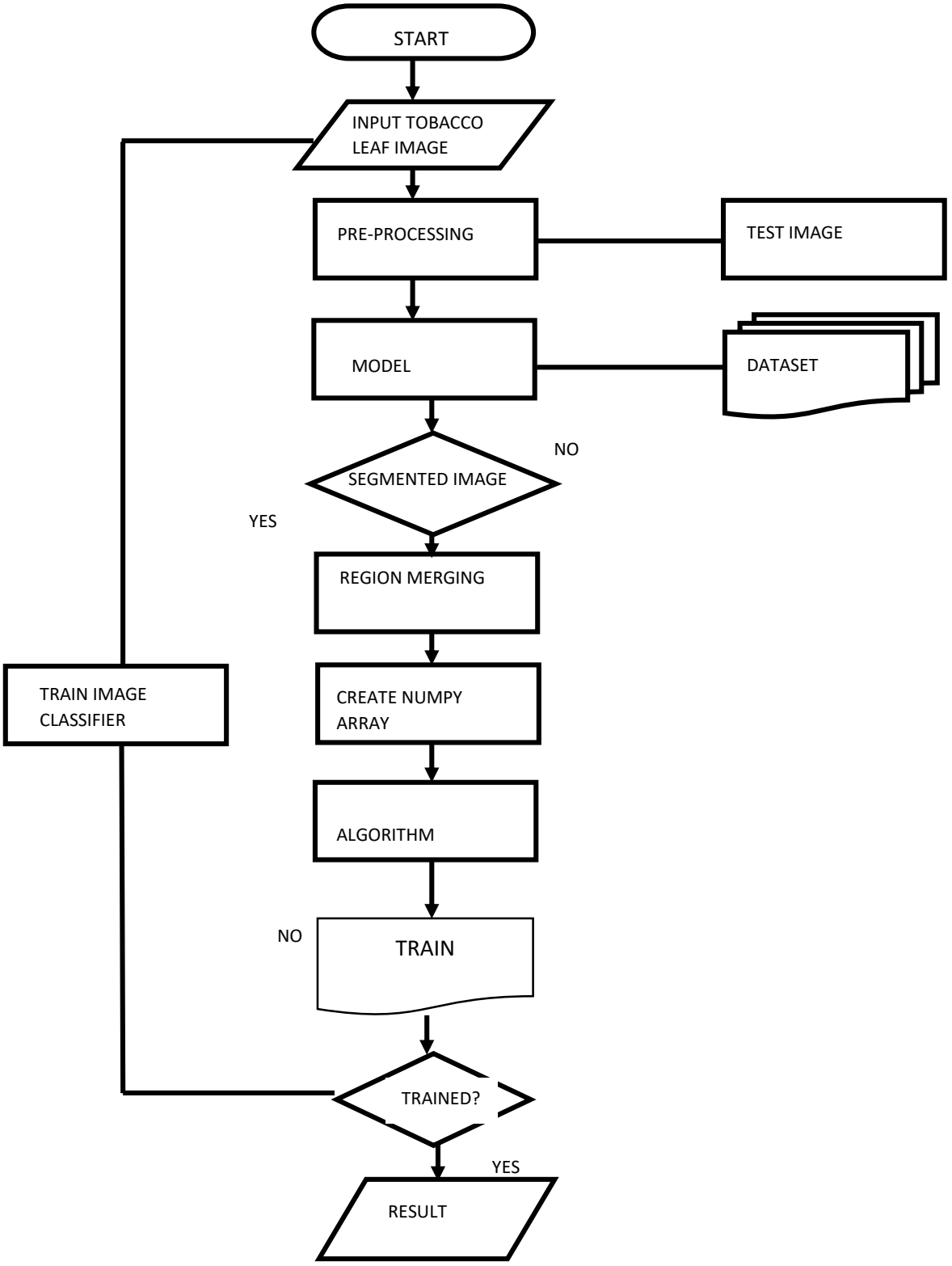

*Figure 3: Flow chart of proposed model*

## 3.8 Data-set

The data-set used for the project consists of images of tobacco leaves for classification purposes. The data-set has been organized into folders based on the quality or grade of the tobacco leaves, utilizing the provided train.csv file. The data-set includes five categories of colored tobacco leaf images: 1805 images of high-grade leaves, 999 images of medium-grade leaves, 370 images of low-grade leaves, 295 images of poor-quality leaves, and 193 images of severely damaged leaves. The distribution of these images across the categories can be visualized using a bar graph, as depicted in the figure below.

## 3.9 Solution

This section shows the solution model, how it predicts the results using the algorithms of deep learning. To come up with a functional solution that would solve the research problem, the researcher made analysis the algorithm to determine the error rate and accuracy rate so as to get more accurate results in classification.

### 3.9.1 Image Processing

#### 3.9.1.1 Resize Image

When defining the architecture of the model, one of the requirements is to define a fixed input form. When performing this task, it is important to keep in mind that there is a balance between speed of computation and loss of information. For clarification, when the size of an image is reduced, information (pixels) is removed. Less information means faster training times; however, it can also mean reduced overall accuracy. an image size of 224 x 224 has been selected.

#### 3.9.1.2 Image Cropping

The tobacco leaf image region will be cropped automatically from each image to remove the background and unwanted region.

#### 3.9.1.3 Gaussian Blur

It is important to address the presence of noise, which can be introduced during the image capture process. Image smoothing techniques play a crucial role in reducing such noise. OpenCV offers various methods for image smoothing, including the use of Gaussian filters. Gaussian filters are preferred due to their ability to reduce noise while preserving important

image features and edges. The image smoothing process was applied using the cv.GaussianBlur() function in OpenCV. A Gaussian kernel with a specified width and height (both positive and odd) and standard deviation values in the X and Y directions (sigmaX and sigmaY) was used. If only sigmaX was specified, sigmaY was set to the same value. By applying Gaussian blur with appropriate parameters, the sharp edges in the tobacco leaf images were smoothed without excessive blurring, effectively reducing Gaussian noise.

To further enhance the quality of the images and mitigate the impact of lighting conditions, additional preprocessing steps were taken. Masks were applied to the images to address lighting variations. The images were then re-sized to a dimension of 224 × 224, ensuring consistency for further analysis. Cropping was performed to remove non-informative areas, retaining only the relevant portions of the tobacco leaves. To obtain the mask, the grayscale conversion of the image was performed, followed by setting a tolerance value greater than 7. This step aimed to eliminate black portions from the image, focusing solely on the informative content. After cropping and re-sizing the image to the desired dimensions, Gaussian blur was applied with a standard deviation of 10 in both the X and Y directions. The Gaussian kernel convolved each point of the input array, resulting in an enhanced output array. The Implementation of gaussian blur aimed to reduce noise, address lighting variations, and enhance the overall quality of the tobacco leaf images for subsequent analysis and classification.

#### 3.9.1.4 Data Augmentation

Data augmentation techniques are employed to expand the size of a data-set by applying transformations to existing examples. This process artificially increases the amount of training data available and can improve the generalization and regularization of machine learning models, as supported by statistical learning theory. Data augmentation has long been recognized as a critical component in machine learning models and has been extensively utilized in various applications. It has been employed to address the common challenge of unbalanced group sizes in datasets. During model training, the objective is to enhance precision in subsequent iterations or epochs. However, under-represented groups may receive less exposure and, consequently, may not be learned as effectively as their over-represented counterparts. Data augmentation helps mitigate the impact of over/under-representation by introducing random changes to the original training images through parameter adjustments.

To achieve this, specific random changes are applied to each image during every epoch of training. These changes can include random rotations ranging from 0 to 90 degrees, random horizontal and vertical flips, and random shifts in both horizontal and vertical directions. By incorporating these random variations into the training process, the model encounters "different" images at each iteration, thereby enhancing its ability to generalize and improve its performance. The original pre-processed images were used for initial network training. Subsequently, real-time data augmentation techniques were implemented throughout the training process to further enhance the network's ability to localize and extract meaningful features. This involved applying random rotations, horizontal and vertical flips, as well as shifts to each image during every epoch. By introducing these random transformations, the model was exposed to a more diverse range of image variations, which aided in improving its overall performance and robustness.

#### 3.9.1.5 Dropout Regularization

Dropout is a technique used during training in neural networks where a random selection of neurons is excluded or "dropped out". This means that these neurons are temporarily ignored during the forward pass, and their contribution to the activation of downstream neurons is disregarded. Consequently, any weight updates are not applied to the dropped-out neurons during the backward pass. The purpose of dropout is to prevent over-reliance on specific neurons and encourage the network to learn more robust and generalization features. As a neural network learns, the weights of neurons settle into their respective roles within the network. Neurons become specialized in detecting specific features, and neighboring neurons start to rely on this specialization. However, excessive specialization can lead to a model that is too tightly bound to the training data, making it fragile and prone to over-fitting.

### 3.9.2 Training Model

In order to develop an effective model for the classification and grading of tobacco leaves, a training process is essential. The training phase involves iteratively adjusting the parameters of the model to minimize the difference between the predicted outputs and the actual labels of the training data. This process enables the model to learn and generalize patterns from the provided data-set. The training of the model consists of several key steps. First, the dataset of tobacco leaf images, along with their corresponding grades or classifications, is prepared. This data-set serves as the foundation for training the model and is crucial for its learning process.

The training process typically involves multiple iterations or epochs, where the model is exposed to the training data-set in batches. During each epoch, the model processes a batch of images, makes classifications, compares them to the actual labels, and adjusts its parameters using optimization algorithms like gradient descent to minimize the classification errors. Throughout the training process, metrics such as accuracy, loss, and validation performance are monitored to assess the model's progress and make informed decisions about its performance. Regular evaluation and validation are crucial to ensure that the model is learning effectively and improving its classification and grading capabilities. The duration of the training process can vary depending on the complexity of the model, the size of the data-set, and the available computational resources. It is important to strike a balance between training for a sufficient number of epochs to allow the model to converge and avoiding over-fitting by stopping the training at the appropriate time. Once the training phase is completed, the trained model is ready to be used for classifying and grading tobacco leaves. It has learned the underlying patterns and features from the training data-set and can make classifications on new, unseen samples with a certain level of accuracy and reliability.

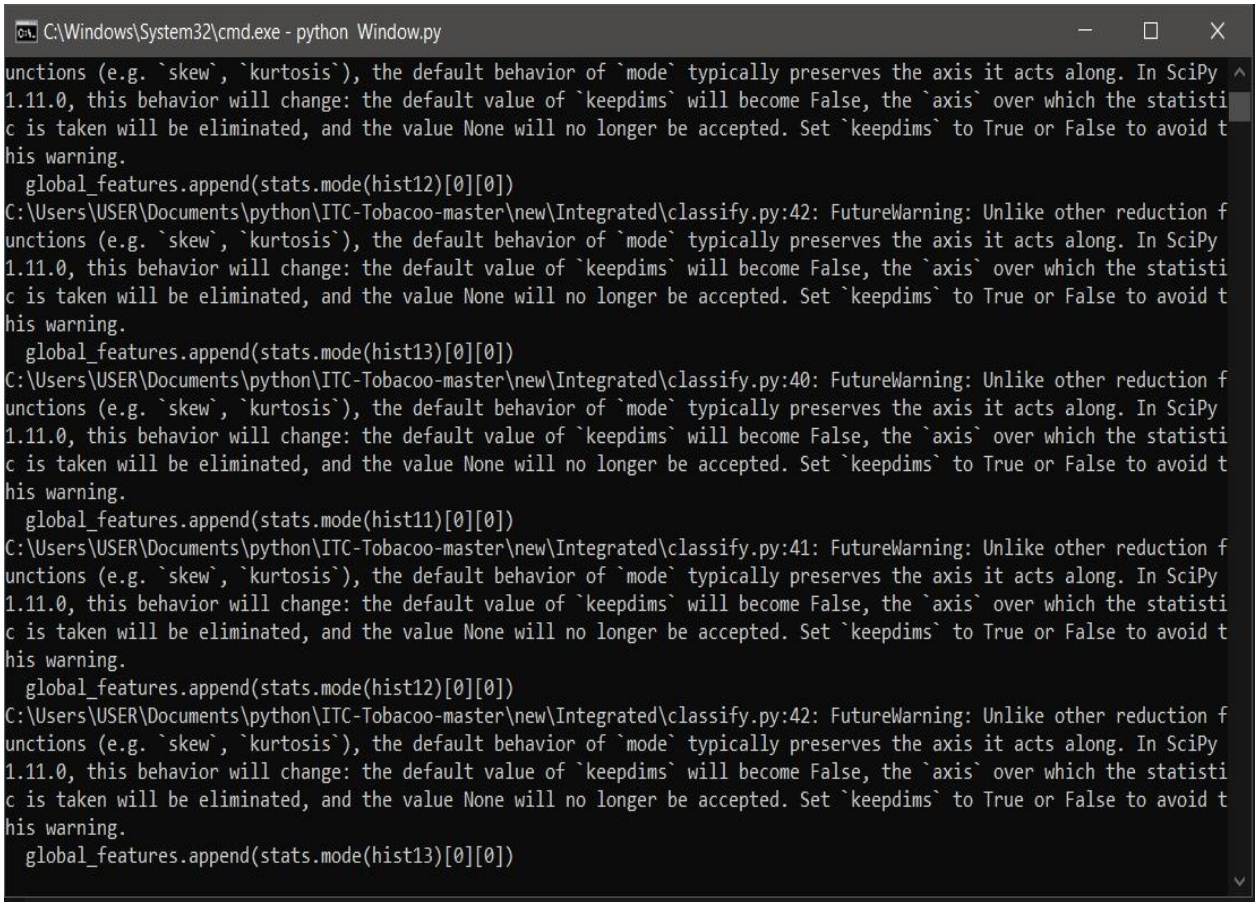


*Figure 4:Saving Model*

## 3.9.3 Evaluating and implementing of Model

## 3.9.3 Implementation

In the implementation phase of our project, we focused on translating the designed model and the trained parameters into executable code. Our goal was to develop a practical software solution that could accurately classify and grade tobacco leaves based on their quality. To begin the implementation process, we set up the development environment by installing the necessary libraries, frameworks, and tools. We chose to use Python as the programming language and leveraged deep learning frameworks such as TensorFlow and PyTorch to facilitate the implementation. Next, we translated the model architecture into code. This involved defining the appropriate neural network layers, specifying their configurations, and establishing the necessary connections according to our designed model. We ensured that the code accurately represented the structure and behavior of our model. We then loaded the trained parameters obtained from the training phase into the implementation code. These parameters, including the weights and biases of the neural network, carried the learned knowledge and enabled our model to make accurate classifications.

Handling input data was an essential part of the implementation process. We developed mechanisms within the code to load and pre-process the tobacco leaf images. This involved tasks such as re-sizing the images, normalizing pixel values, and applying any required transformations to ensure consistency with the training process. With the model architecture, trained parameters, and input data handling in place, we implemented the functionality to classify and grade tobacco leaves. By passing the pre-processed input images through our model, we obtained classifications or scores for each leaf. The implementation phase of our project was successful in translating the designed model and trained parameters into executable code. We developed a functional software solution capable of accurately classifying and grading tobacco leaves based on their quality.

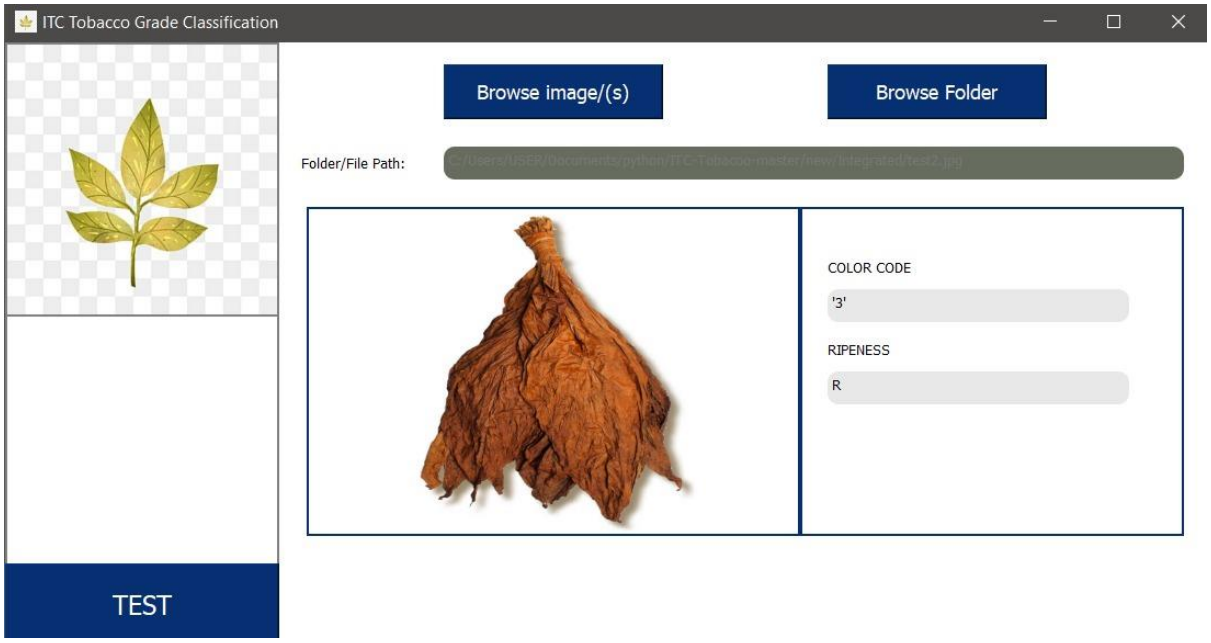


*Figure 5:Snapshot of model implementation*

## 3.10 Summary

In this chapter, we focused on the development of our model, discussing the methods and tools utilized throughout the process. The model was developed using Python and neural network frameworks, tailored specifically for our project. Prior to training, we applied techniques such as converting the data to Gaussian blur and implementing dropout regularization to prevent over-fitting. To facilitate the development process, we utilized the Python Jupyter Notebook as our integrated development environment (IDE). By employing a prototyping model for system development, we were able to effectively utilize the various processes involved in completing the model within the designated time-frame. The training phase resulted in a satisfactory accuracy rate of 90.4%. In the upcoming chapter, we will delve into additional performance measures, including specificity, recall, AUC, ROC, F1 measure/F1 score, and precision, to further evaluate the effectiveness and reliability of our model. This chapter outlines the development stages of our model, emphasizing the techniques, tools, and approaches employed to achieve our objectives. The subsequent chapter will provide a comprehensive analysis of the model's performance, offering a deeper understanding of its capabilities and limitations.

# CHAPTER 4: RESULTS

## 4.0 Introduction

After completing the development of the tobacco leaf grade classification system, it was essential to evaluate its effectiveness and efficiency. Measures were taken to assess the system's performance, accuracy, and response time. The analysis of the previously collected data was instrumental in deriving meaningful results. Various testing techniques, including white box, black box, and unit testing, were employed to observe the system's behavior under different circumstances. The evaluation process aimed to determine the system's ability to accurately classify tobacco leaves and its efficiency in terms of response time. Accuracy served as a performance metric to measure the system's capacity to correctly classify tobacco leaves. Additionally, response time was evaluated to gauge the system's speed and responsiveness in processing and delivering results.

To ensure the system's reliability and robustness, a range of testing approaches was implemented. White box testing involved examining the internal components and logic of the system to identify any potential errors or flaws. Black box testing focused on evaluating the system's outputs and functionality without delving into its internal workings. Unit testing was conducted on individual components or modules to verify their accuracy and ensure proper integration within the system. By utilizing a combination of these testing techniques, the evaluation process provided valuable insights into the system's performance, accuracy, and response time. It helped identify areas that required improvement and ensured that the developed system solution met the specific objectives and requirements of the tobacco leaf grade classification project.

## 4.1 System Testing

Testing is a crucial aspect of system development as it ensures the reliability and effectiveness of the developed system. This chapter presents the conducted tests and the corresponding outcomes, with a particular emphasis on examining both functional and non-functional requirements of the proposed solution.

### 4.1.1 Black Box Testing

Black box testing is a method where the internal workings, structure, and implementation details of the product are not considered. In other words, the tester is unaware of the system's internal operations. Black box testing focuses solely on evaluating the system's external behavior. It involves testing the system's inputs and analyzing its outputs or responses. The

outcomes of the black box tests conducted on the tobacco leaf grade classification model are as follows. The system will be assessed to determine its accuracy in classifying the features of tobacco leaves. The subsequent section presents the results obtained from the author's black box testing of the model.

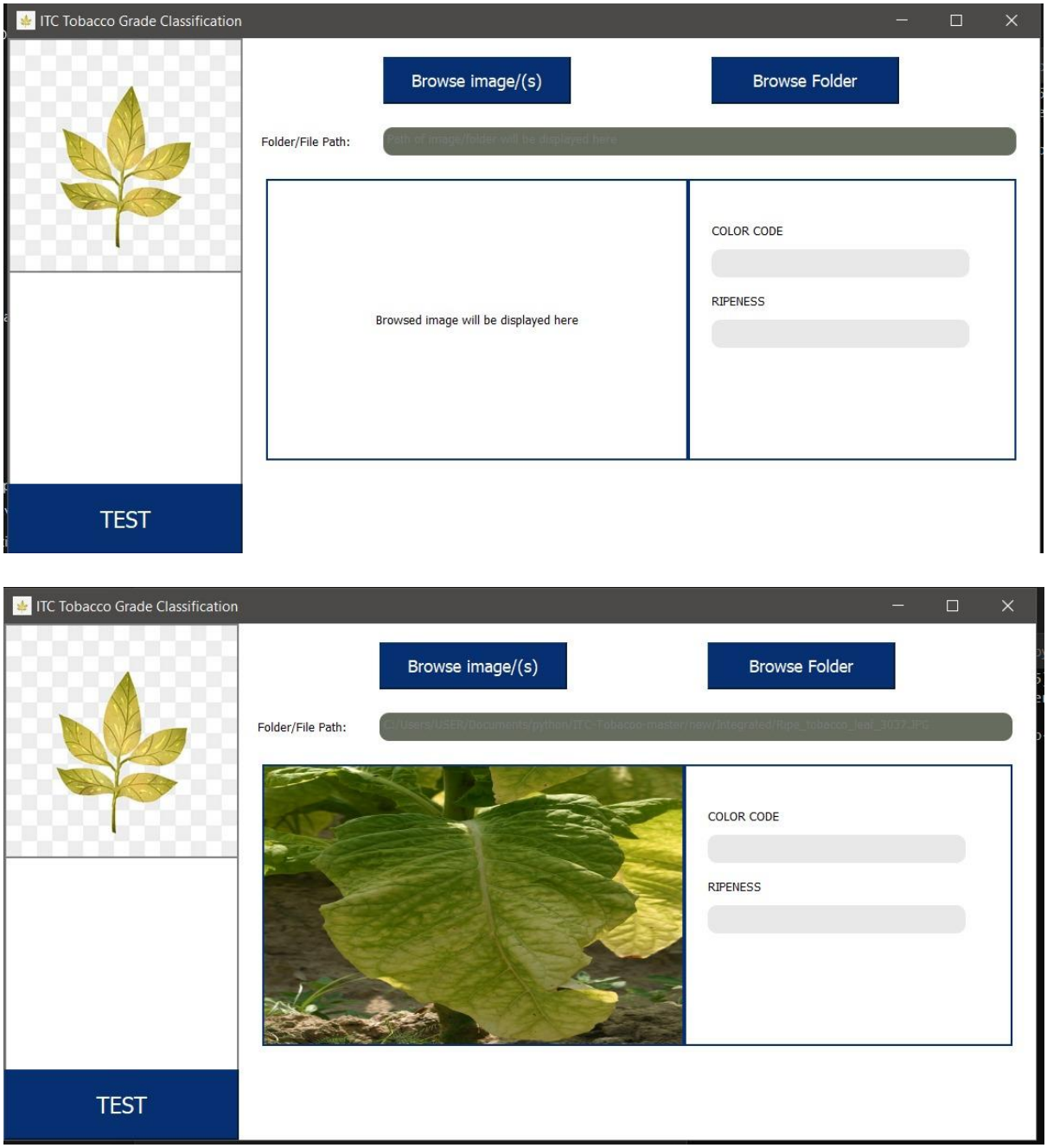


*Figure 6: Testing system input and output*

## 4.1.2 White Box Testing

White box testing is a software testing approach where the tester possesses knowledge of the internal structure of the software before conducting the tests. This type of testing is usually

performed by software developers who have a deep understanding of programming and implementation. White box testing is applicable to lower levels of testing, such as unit and integration testing. Its main focus is to examine the computer code of the system being tested, including its branches, conditions, loops, and overall code structure. The primary goal of white box testing is to assess the functionality of the system. The developer conducted tests on the model, as demonstrated below:

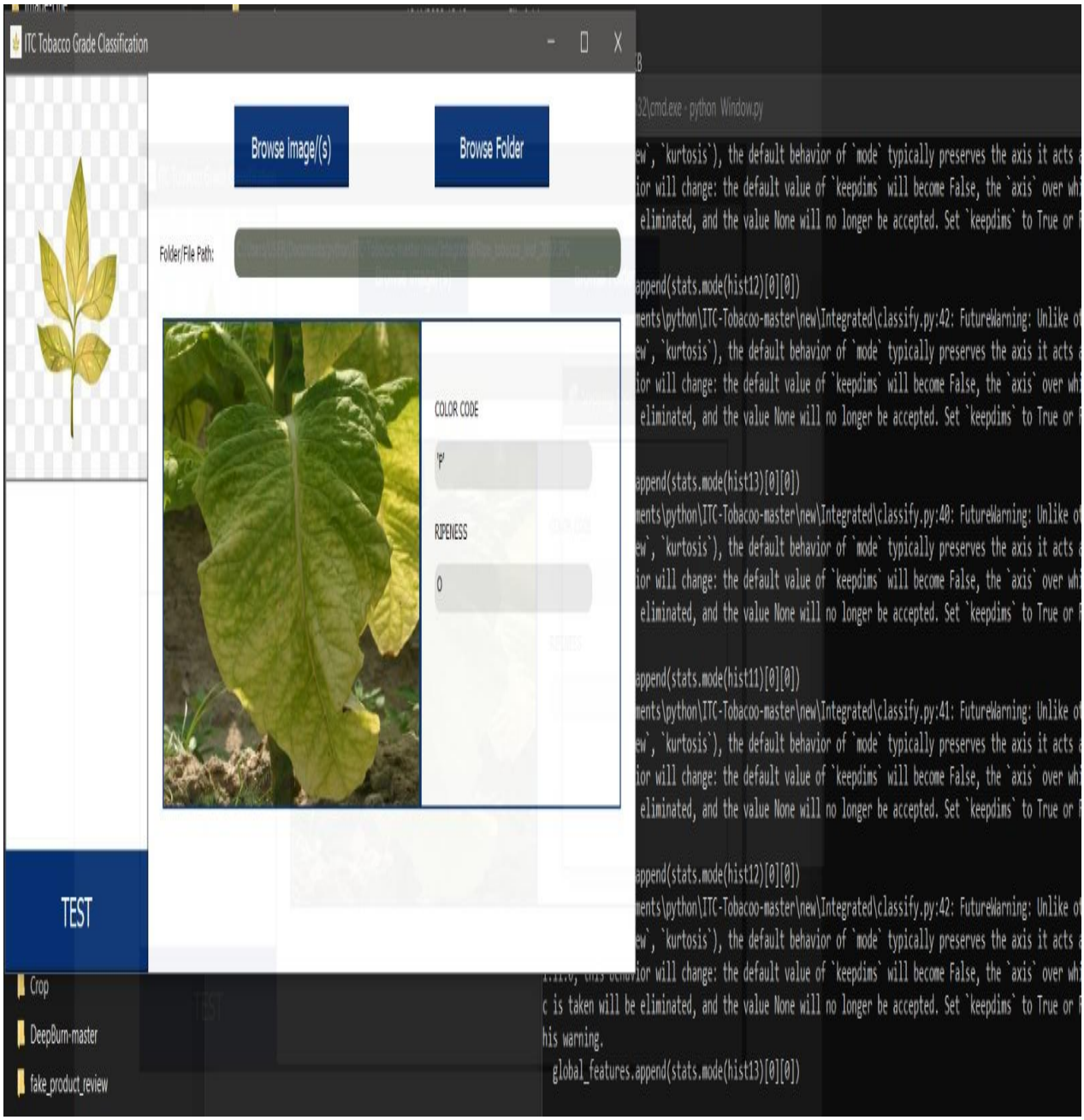


*Figure 7: testing system code*

## 4.2 Evaluation Measures and Results

The evaluation of a model's effectiveness involves the utilization of performance evaluation metrics (Hossin & Sulaiman, 2015). As stated by Hossin & Sulaiman (2015), model assessment metrics can be categorized into three types: threshold, probability, and ranking. In this project, the performance of the system is assessed based on its ability to accurately classify tobacco leaves. To evaluate the accuracy of the system, the author employed a confusion matrix.

### 4.2.1 Confusion Matrix

The confusion matrix is a tabular representation that displays the number of instances classified correctly and incorrectly by a model. It is used to assess the performance of the model. The confusion matrix consists of four terminologies: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP refers to instances that are truly positive and are correctly classified as positive by the model. TN represents instances that are truly negative and are accurately classified as negative by the model. FP represents instances that are actually negative but are mistakenly classified as positive by the model. FN represents instances that are truly positive but are incorrectly classified as negative by the model. These terminologies help evaluate the accuracy and effectiveness of the model in classifying instances in the tobacco leaf grade classification project.

| Type | Returned number of correct classifications | Returned number of incorrect classifications |
|---|---|---|
| 1 | True Positive | False Negative |
| 2 | False Positive | True Negative |

Table 1 Confusion Matrix

The technology was put to the test in terms the returned number of correct and incorrect classification. For the purpose of observing the system's findings, three scenarios and a test environment were developed. The system was observed 40 times on each scenario using different testing input. All of the scene analysis was done to ensure that the answer was accurate and that false classifications were identified. The tables below indicate the outcomes of the tests that were conducted.

**Table 2 Confusion matrix for tobacco leaf classification**

| Test cases | classifications | Number of tests | Correct classification | False classification | Classification |
|---|---|---|---|---|---|
| 1 | Yes | 40 | 40 | 0 | True positive |
| 2 | No | 40 | 38 | 2 | True negative |

Table 2

## 4.3 Accuracy

The number of correct classifications divided by the total number of tests in each category equals accuracy. The percentage of accuracy is then calculated by multiplying it by 100. The following equation is used to compute it:

Equation 1: Accuracy calculation as adopted from Karl Pearson (1904)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100$$

$$\text{Accuracy rate} = \frac{40 + 38}{40 + 38 + 0 + 2} * 100$$

$$\text{Accuracy} = \frac{72}{80} * 100$$

$$\text{Accuracy} = 98{,}5\%$$

## 4.4 Misclassification Rate/ Error Rate

- Overall, how often is it wrong?
- It tells you what fraction of classifications were incorrect. It is also known as Classification Error.
- **Error rate = (FP+FN)/(TP+TN+FP+FN) or (1-Accuracy**

  = (0+2)/( 40+38+0+2)

  **=0.025%**

## 4.5 Precision

- When it predicts yes, how often is it correct?

**Precision = TP/(TP+FP)**

40/(40+0)

**=100%**

## 4.6 Sensitivity/Recall/True Positive Rate

- When it's actually yes, how often does it predict yes?
- It tells what fraction of all positive samples were correctly predicted as positive by the classifier. It is also known as True Positive Rate (TPR), Sensitivity, Probability of Detection.

**Recall = TP/(TP+FN)**

**=40(40+2)**

**=95%**

## 4.7 Specificity/True Negative Rate

- When it's actually no, how often does it predict no?
- It tells what fraction of all negative samples are correctly predicted as negative by the classifier. It is also known as True Negative Rate (TNR).
- Equivalent to 1 minus False Positive Rate
- **Specificity = TN/(TN+FP) or 1-FP rate**

=1-2

**=98%**

## 4.8 Prevalence

- How often does the yes condition actually occur in our sample?
- It shows how often does the yes condition actually occur in our sample
- **Prevalence=Actual YES/(TP+TN+FP+FN)**

## 4.9 F1-Score/F1 Measure

- It combines precision and recall into a single measure.
- **F1-score=2 x (Precision x Recall/ Precision + Recall)**

**=2TP/(2TP+FP+FN)**

**=2(40)/(2 x 40 + 0+2)**

**=97.5%**

## 4.5 Summary of Research Findings

The researcher performed all the necessary black, white box tests and performance tests using the confusion matrix, the author found that the system had satisfactory performance. The system was tested in accuracy, misclassification error/error rate and it achieved 98.5% and 0.025% respectively. The model attained an overall precision of 100% and a sensitivity or recall of 95%. An F1 score of 97.5% was achieved with a specificity or true negative rate of 98%.

## 4.6 Conclusion

To conclude this chapter, the author used different metrics for performance measurement of the system. Among them being, accuracy, specificity, recall precision, error rate f1-score and true positive rate. The next chapter presents the conclusion, objective realization and recommendations for further development.

# Chapter 5: Conclusion and Recommendations

## 5.1 Introduction

This chapter marks the conclusion of the research and provides a reflective analysis to determine the achievement of the study's objectives. It presents a concise overview of the findings, summarizes the conclusions drawn from the research, and offers recommendations for future studies.

## 5.2Aims & Objectives Realization

The first objective of this study was to analyse different machine learning techniques used for tobacco leaf classification. The second objective was to design and implement a machine learning model which classifies tobacco leaves. The third and last objective was to evaluate the effectiveness of machine learning learning model in tobacco grading and classification. Therefore, to this end, the researcher developed a model that using a convolutional neural network (CNN) to classify the features of a tobacco leaf and predict its ripeness which satisfies the second research objective. The researcher performed all the necessary black, white box tests and performance tests using the confusion matrix, the author found that the system had satisfactory performance. The system was tested in accuracy, misclassification error/error rate and it achieved 98.5% and 0.025% respectively. The model attained an overall precision of 100% and a sensitivity or recall of 95%. An F1 score of 97.5% was achieved with a specificity or true negative rate of 98%. A validation accuracy of 95%, mean percentage error of -0.044% were achieved. Therefore, providing an improvement over previously implemented similar solutions. The model is easily available as it is implemeted on a mobile phone.

## 5.3Major Conclusions Drawn

Firstly, it is evident that the current grading process, which relies on subjective evaluation by human graders, poses a significant problem for farmers in Zimbabwe. This subjectivity opens the door for potential exploitation, as farmers may unknowingly sell their tobacco bales at lower prices than they deserve. This highlights the need for an objective and accurate grading system that can empower farmers to negotiate fair prices for their produce. Secondly, the developed machine learning model for tobacco leaf grade classification demonstrates its effectiveness in accurately assessing and classifying tobacco leaves. By utilizing deep learning techniques, specifically the Random Forest algorithm, the model achieves high accuracy in its predictions. This suggests that the model can reliably distinguish between different grades of tobacco leaves based on learned patterns and features. Furthermore, the evaluation results of the model indicate its satisfactory performance across various performance metrics. The model achieves high accuracy, with a low misclassification error rate. It also demonstrates excellent precision, recall, and F1 score, indicating its ability to make precise predictions while maintaining a good balance between identifying true positives and minimizing false positives.

The findings of this research confirm the significance and efficiency of using deep learning architectures, particularly LSTM, in predicting tobacco leaf grades accurately. The developed model offers an automated and consistent assessment of tobacco leaf grades, enabling farmers to have a fair understanding of the value of their produce. By empowering farmers with objective grading information, the model contributes to creating a more transparent and equitable trade environment in the tobacco industry.

## 5.3 Recommendations & Future Work

Firstly, it is recommended to further enhance the developed machine learning model by incorporating additional features and data sources. For instance, considering factors such as leaf texture, color, and size could potentially improve the accuracy of the classification. Additionally, integrating real-time environmental data, such as temperature and humidity, could provide valuable insights into the impact of growing conditions on tobacco leaf grades. In addition, expanding the dataset used for training the model would be beneficial. Increasing the diversity and volume of the dataset can help capture a broader range of variations in tobacco leaf characteristics and improve the generalization ability of the model. Collaborating with tobacco farmers and industry stakeholders to gather more comprehensive and representative data would contribute to the robustness of the model.

Furthermore, conducting field tests and validations of the developed model in real-world tobacco farms is recommended. This would provide an opportunity to assess the model's performance under different environmental conditions and farming practices. Collecting feedback from farmers and incorporating their insights into the model's development can further enhance its practical applicability and user satisfaction. Additionally, exploring the feasibility of integrating the developed model into existing tobacco grading systems is worth considering. Collaborating with industry partners to implement the model in commercial grading processes could streamline the grading workflow and provide real-time grading results. This integration would require addressing technical and logistical challenges and ensuring compatibility with existing infrastructure.

## References


1. Nyabadza, N. (2019). Farmers bemoan tobacco grading system. The Herald. Retrieved from https://www.herald.co.zw/farmers-bemoan-tobacco-grading-system/
2. Munyaradzi, T. (2020). Farmers want transparency in tobacco grading. The Standard. Retrieved from https://www.thestandard.co.zw/2020/03/01/farmers-want-transparency-tobacco-grading/
3. Tobacco Reporter. (2023). High Tobacco Auction Rejection Rate. Tobacco Reporter. Retrieved from https://tobaccoreporter.com/2023/03/23/zimbabwe-tobacco-auction-rejection-rate-high/

4. Akram MU, Khalid S, Tariq A, Khan SA, Azam F (2014) Detection and classification of tobacco leaves. Comput Biol Med 45:161–171. https://doi.org/10.1016/j.compbiomed.2013.11.014
5. H. Pratt, F. Coenen, D. Broadbent, S. Harding, Y. Zheng, "Convolutional Neural Networks for Tobacco Classification", International Conference on Medical Imaging Understanding and Analysis, Loughborough, UK, July 2016.
6. Cheng X, Wong DWK, Liu J, Lee BH, Tan NM, Zhang J, Cheng CY, Cheung G, Wong TY (2012) Automatic localization of tobacco features. Ann Int Conf IEEE Eng Med Biol Soc 2012:4954–4957. https://doi.org/10.1109/EMBC.2012.6347104
7. Gutiérrez, J., Vargas, J., Rojas, C., & Solorio, T. (2019). A Convolutional Neural Network-based Approach for Tobacco Leaf Classification. 2019 International Joint Conference on Neural Networks (IJCNN), 1-8.
8. World Health Organization. (2022). Tobacco. Retrieved from https://www.who.int/news-room/questions-and-answers/item/tobacco
9. Singh, S., & Yadav, A. (2020). Tobacco leaf grading using deep learning. 2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 1-4.
10. Malik, S., Raheem, A., & Sharif, M. (2021). Tobacco leaf grading using an ensemble of classifiers. Neural Computing and Applications, 1-12.
11. Bora, N., Singh, A. K., & Das, D. (2020). A robust tobacco leaf grading system based on color and texture features. Computers and Electronics in Agriculture, 174, 105470.
12. Dey, D., Das, S., & Mukherjee, J. (2021). A Hybrid Approach for Tobacco Leaf Grading Using Deep Learning and Color-Texture Features. Journal of Imaging, 7(6), 99.
13. Ma, Y., Zeng, M., Wang, Y., & Zhang, L. (2021). Research on tobacco leaf grading system based on machine learning. Journal of Ambient Intelligence and Humanized Computing, 12(4), 3961-3970.
14. Sahoo, S., Nayak, R., Mishra, S., & Nayak, A. (2020). Tobacco leaf grading using deep learning-based CNN approach. Journal of Ambient Intelligence and Humanized Computing, 11(11), 5091-5098.
15. Government of Zimbabwe. (2021). Tobacco. Retrieved from https://www.zimtreasury.gov.zw/sectors/tobacco/

16. Tobacco Industry and Marketing Board. (n.d.). About TIMB. Retrieved from http://www.timb.co.zw/index.php/about-timb
17. Zimbabwe Tobacco Association. (n.d.). Who we are. Retrieved from https://zta.co.zw/about-us/
18. Tobacco Research Board. (n.d.). About TRB. Retrieved from http://www.trb.co.zw/about.html
19. The Herald. (2019, April 3). Tobacco farmers lose millions through grading fraud. The Herald. Retrieved from https://www.herald.co.zw/tobacco-farmers-lose-millions-through-grading-fraud/
20. Zimbabwe Independent. (2019, April 26). Farmers losing out on tobacco sales. Zimbabwe Independent. Retrieved from https://www.theindependent.co.zw/2019/04/26/farmers-losing-out-on-tobacco-sales/