BSCS-S21-001

03-134181-008 JIBRAN ALI

03-134181-021 MUZAMIL RASHID

Al-Biruni Gemstones Authenticator

In partial fulfilment of the requirements for the degree of

Bachelor of Science in Computer Science

Supervisor: DR. MUHAMMAD AASIM QURESHI

Department of Computer Sciences

Bahria University, Lahore Campus

December 2021

C e r t i f i c a t e

We accept the work contained in the report titled.

"Al-Biruni Gemstones Authenticator"

written by

JIBRAN ALI
MUZAMIL RASHID

as a confirmation to the required standard for the partial fulfilment of the degree of Bachelor of Science in Computer Science.

Approved by:

Supervisor: DR. MUHAMMAD AASIM QURESHI

_____

December 30, 2021

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Enrolment	Name	Signature
03-134181-008	JIBRAN ALI	
03-134181-021	MUZAMIL RASHID	

Date : December 30, 2021

Specially dedicated to
my beloved Parents, teachers
(JIBRAN ALI)
my beloved Parents, teachers
(MUZAMIL RASHID)

ACKNOWLEDGEMENTS

Thanks to almighty ALLAH, my dear parent and our supervisor, Dr. Muhammad Aasim Qureshi, for his patience, guidance, and support. We have benefited greatly from your wealth of knowledge. We are extremely grateful that he took us on as students and continued to have faith in us over the time.

In addition, we would like to thank our friends and family who supported us and offered deep insight into the study.

JIBRAN ALI
MUZAMIL RASHID

AL-Biruni Gemstone Authenticator

ABSTRACT

A gemstone is a mineral that is produced from geological processes. All gemstones are either sorted by their colour, nature, or hardness. Gemstones have the potential to make improvement in the export earnings of Pakistan. Gemstones mines inside Pakistan can generate 4-5 or more billion dollars. The complexity of the product is high, audience often suffer loss. It is very difficult to tell difference between real and fake gemstones. The laboratory identification test is time consuming and expensive. There is a need of fast and cheap expert gemstone identification system.

The main objective is to create a system that would be able to identify multiple types of gemstone types, namely {Turquoise, Ruby, and Emerald}. AL-biruni Gemstone is gemstones identifier based on machine learning. Machine learning is based on the idea that system can learn from data and forecast prediction on unseen data. Machine learning includes series of learning techniques—supervised, unsupervised, reinforcement learning. The machine learning technique used for design of model is convolution neural network.. The benchmark of the system is to create a system having accuracy more than 90%.with large dataset The model has 12 layers.. The hyper parameter tuned is epoch, batch size and learning rate. The learning rate used were 0.001, 0.0001, and 0.00025. The batch size were 8, 12, and 16.

The Application is developed as prove of concept named Gemo. Gemo android Application is used for integration with machine learning model. The application provides two main features—Authentication and Identification. The authentication features take image as input and return gemstones type as output. The Accuracy of the model used in the application for authentication is 95.0%.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Geology is one of the interesting yet to be more explored field of the science. Most unrecognized branch of geology is Gemmology. This branch has its unique importance. A gemstone is a mineral that is produced from geological processes. According to the Mental fuss there are more than 200 types of gemstones in the world. All gemstones are either sorted by their colour, nature, or hardness. The hardness is measured upon Mohs Scale. Mohs scale was created by Fredrich Mohs in 1812. The scale is of range from 0 to 12. Usually, Gemstones have hardness greater than 7 Mohs. Greater the gemstones hardness, more precious and valuable they are.

Al-Biruni was great Muslim scientist who added remarkable contribution to mineralogy. Gemstone can be powerful factor to increase the country financial side. Many countries like India, US, Hong Kong, and Thailand uses Gemstones as their trump card for the export earnings. Muslims have rich history with gemstones. Studies shows that numerous Muslim emperors used to wear rings with gemstones in it. Tipu sultan was huge admirer of the gemstones. As the gemstone subject is getting popular again in 21st century but very few findings had been contributed to the gemmology by south-Asian writers and researchers.

1.2 Motivation

Pakistan is gifted with gemstone's mines. Especially area around **Swat** is filled with *EMERALD, RUBY, AND SPINEL*. Majority of people having interest in gemstones are unable to identify and authenticate them. Because it is not an easy job to remember all types of the stones. Majority remember them by colour, some by their shines etc. However, the world of gemstones is large, one's colour is found in many gemstones of different kinds. To differ between them can be complex job for the individual with no knowledge.

Despite being rich with gemstone's resources Pakistan is unable to utilize them. *Gemstone only add 0.31 to export earnings*. Because of lack of knowledge, technology gemstones are illegally sent to USA. USA further works on these precious stone and exports to earn. *According to oec. world USA was the top exporters of the gemstone with earnings of 1.56 billion in 2019*. Pakistan used to have Gemstone Corporation of Pakistan but was shut down in 1997.
Knowledge of gemmology is detailed, not everyone has the time to grip it. The barrier has kept gemstones to not grow both national and international. People are scammed at the national level because of their non-expertise in the gemmology. Gemstones sell to them are either fake, or not of their desired type or sell at high price. Creation of such system that can help common people and government has inspired us to do such project.

1.3 Primary Scope

Pakistan is gifted with these minerals. Gemstones can be next billion-dollar project for Pakistan. Pakistan has mines of gemstones like Emerald, Ruby, and Peridot. After diamond, the most expensive gemstone is emerald. In [1], mines of emerald situated in Mingora, Gujar Killi and Shaozai alone can generate **2-3 billion** US dollars. Remaining Gemstones can also be exported to generate earning around **1-2 billion** US dollars. However, Nothing Noticeable has been done yet. The main reasons are no training institutes in the country, few experts and lack of knowledge. There are

countable number of gemologists in Pakistan. Only few skilled at their job but offering services at high rates.

The details regarding scope are as follows.

1.3.1 Included

Scope of the project is limited to 2 types of gemstones. Gemstone identification and authenticity (real or fake) are the main functionality of the project. Following are the types of the gemstones included.

- Turquoise
- Emerald

1.3.2 Not Included

All the gemstones apart from the two above are not included in the scope of the project. Following are the characteristics of not included gems and reasons for absence.

- Gemstone that are rare. It is difficult to collect data for these stones
- Gemstones that are expensive. Including these gems would increase the budget of the project
- Alone copper-based stones. Difficult to remove lighting factor for these kinds of stones

1.4 Optional Scope

In our optional scope we are aiming to add two more gemstones for identification and authentication. Following are the aimed gemstones.

1. Agate
2. Ruby

1.5 Problem Statement

The proposed solution is trying to solve a common problem audience facing regarding gemstones. Since, the world of gemstones is so vast, to prevent frauds, comprehensive understanding is necessary. The product complexity is high it is ranked 838th. Gemstones was 356th most traded product in 2019. People often get scammed in this subject. The product they buy are usually fake.

To verify the gemstone identity, they need to go through the process of identification using laboratory. Gemstones identified using laboratory often loss their natural glow and Attraction. Identification fee for the gemstone is also high. The fee identification depends upon the type of gemstone. In Pakistan government sponsored laboratory charge 1000 for the identification of gemstone which is lowest in the region. The current solution is expensive and time-consuming.

Problem statement is because of the complexity of the gemstone, identification is hard. Current Solution used of this problem is hard and time consuming.

Proposed solution will authenticate between real and fake gemstones of three kinds—Emerald, Ruby, Turquoise.

1.6 Objective

The objectives of this projects are.

- To identify the type of gemstones
- Creation of benchmark dataset
- To Check the authenticity of the Gem’s stone-real or fake

1.7 Limitation

- The image must contain stones inside the image else accuracy of the model is compromised.

1.8 Final Deliverable of the Project

The final deliverable of our project will an App. This app will take the image of the gemstone as input. This app provides the following features, Types of gemstones can be identified as per its authenticity. This application would use android technology and Neural network model for the identification and authentication purpose. At the end output will be shown back to the user interface.

CHAPTER 2

Software Requirement Specification/Literature Review

2.1 Introduction

SRS stands for software requirement specification. This document will describe the functional and non-functional requirement for the system. The limitation that the application will have and reasons for the limitation. The system main features and blueprint of the system. The scale used for the assigning importance is low, medium, and high. Project is research based. The main category of the project is research but to prove the concept implementation of the application or system is compulsory. The system requirements of Al-Biruni Gemstone Authenticator are discussed below.

2.2 Purpose

The Purpose of this document is to describe the authentication system that will aim to help gemmologist and common people for authentication of gemstones. The second purpose is explaining the requirement expected of the system and their cost.

2.3 Intended Audience

Anyone who wishes to authenticate the sort of gemstones, including gemmologists, gemstone enthusiasts, and others.

2.4 Overall Description

In this section system whole description will be provided. Below is the detail.

2.4.1 Product Perspective

A system stores the following information.

1. Dataset*:*

Datasets have images of real gemstones and fake gemstones to identify and differentiate between gemstones.

2. Machine Learning:

A machine learning algorithm to identify and differentiate between gemstones.

2.4.2 User Classes and Characteristics

There is only one user class. The name of that class is gemmologist.

Gemologist – Gemologist can perform the following actions:

- Login
- Sign Up
- Forgot Password
- Reset Password
- Take Picture

- Upload Image
- Edit Info
- Change Image
- View the Identification Result
- Save the Results

2.4.3 Product Features

We will provide an application to identify the types of gemstones and it will also authenticate the originality of stance. The major two features are identification and authentication. The ERD of the system is shown below as Figure 2-1. System ERD

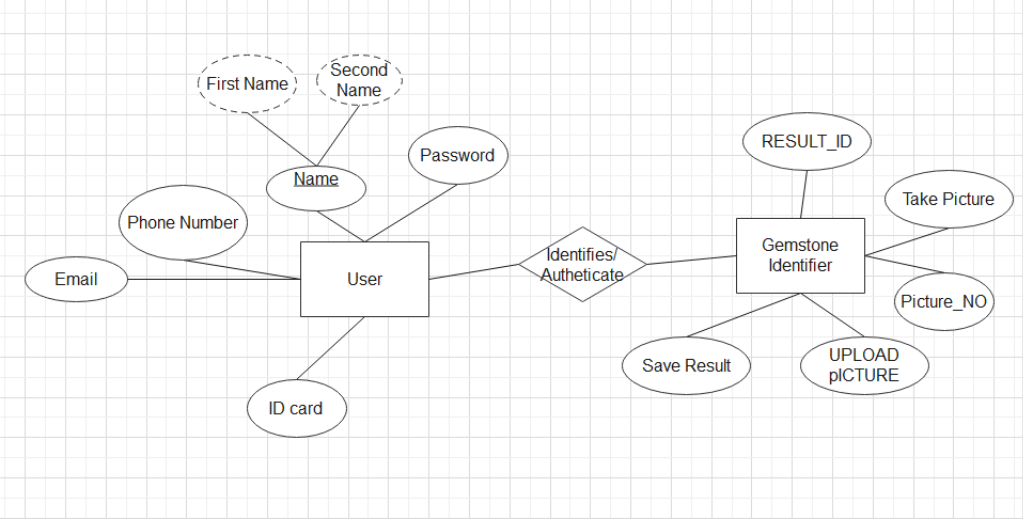


Figure 2-1. System ERD

User must enter name, email, or phone to register. There are two entity user and identifier. Identifier would identify and authenticate the result back to the screen. It will differentiate between real and fake gemstones. We are using only 2 stones, but it will help them. As the emerald and turquois are very expensive gemstones.

2.4.4 Operating Environment

Operating environment of Al-Biruni Gemstone Authentication system is as listed below:

- Operating system: Android
- Software Platform: Android version 24 or more
- Hardware: Android hardware with 56-64 mb
- Co-exist: Android Emulator

2.4.5 Design and Implementation Constraints

Programming standard language used are java for android development and python for deep learning. The databases used is firebase. The issue that can limit the option offered by application is availability of no dataset. The dataset required for the deep learning is not available, we must create our own dataset. Because of Geographical limitation, we cannot capture all type of gemstone. The accuracy of the application will be heavy damaged by the above issues.

The hardware limitation is not all android phones with 24 version comes with good camera. Different cameras different accuracy is expected. The processing is also dependent upon the processor of the android.

2.4.6 Assumptions and Dependencies

The proposed project involves risk of no data available. No major help is available on digital platform regarding this project. Dataset should be readable, trainable, and testable by machine. Algorithm which we used in our project could be right one. The labeling of the dataset is an assumption that is directly related to the project's

performance. We're going to assume that the jeweler's gemstone information is correct. If a jeweler considers a certain gemstone to be a fake, we will mark it as such. If any incorrect gemstone information is given, the system's accuracy will be jeopardized.

2.4.7 System Use Cases

The system use cases are discussed below.

2.4.7.1 Signup(U1)

This use-case will register the user into the system.

Table 2-1. Signup Use case

	Name	Signup
1.	**Use-Case ID**	U1
2.	**Objective**	To register the user to use full feature of application
3.	**Priority**	Low
4.	**Source**	Muzamil (Developers)
5.	**Actors**	User
6.	**Flow of Events**	• Run the application. • Select the Signup option. • Enter Name • Enter Email • Enter Password • Enter CNIC • Press the signup button

6.1	**Basic Flow**	A new user is created, and he can use the system features.
6.2	**Alternate Flow(s)**	If information is invalid, error will display, and user can enter new information.
6.3	**Exception Flow(s)**	User should enter correct details else he will not signup
7.	**Includes**	No other use case
8.	**Preconditions**	The User Mobile should be connected to the internet and application must be installed. Application is running.
9.	**Post conditions**	The User is registered, and login screen will appear.
10.	**Notes/Issues**	None

The user will use the signup option and enter the details. The detail will be authenticated and forward to the database for registration. Upon success a message is return of user registered successfully. On error or unsuccessful attempt, error message is shown to the screen.

Graphical representation of the use case is as show in Figure 2-2 Use Case Diagram Signup.

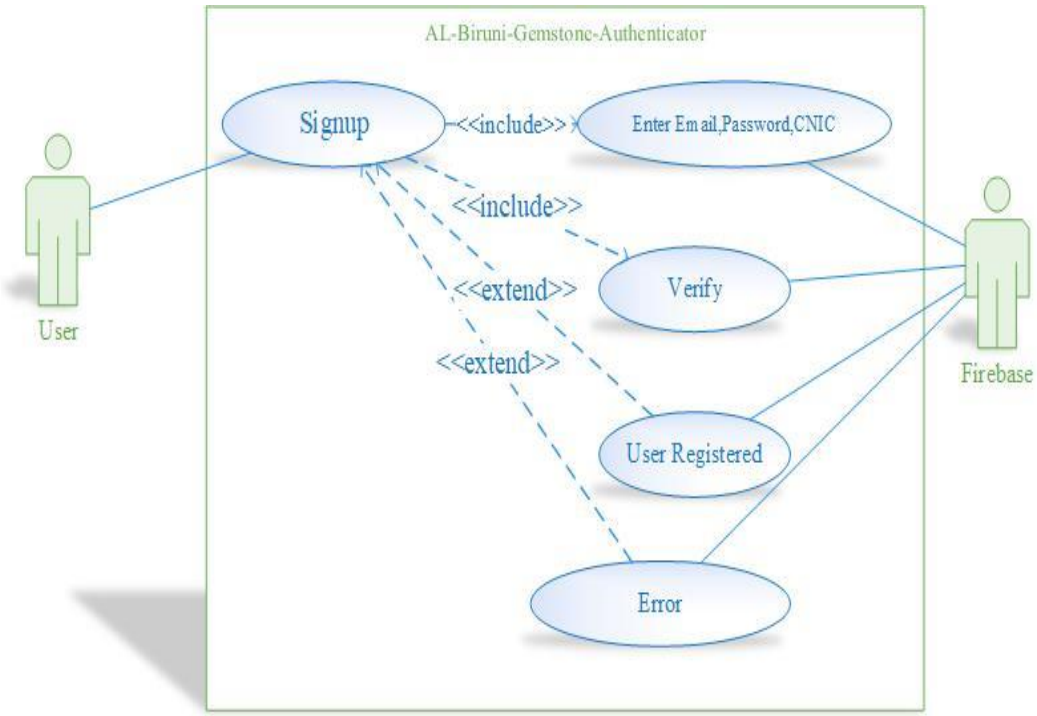


Figure 2-2 Use Case Diagram Signup

2.4.7.2 Login (U2)

This use case would be used to authenticate the user details and registration of the user.

Table 2-2 Login Use Case

	Name	**Login**
1.	**Use-Case ID**	U2
2.	**Objective**	To login to the system
3.	**Priority**	Low
4.	**Source**	Muzamil (Developers)
5.	**Actors**	User

6.	**Flow of Events**	• Run the application. • Select the login option. • Input User_id • Input Password • Press the login button
6.1	**Basic Flow**	The information is entered, and the user can now login
6.2	**Alternate Flow(s)**	If in the basic flow the system cannot find the name or the password is invalid, an error message is displayed. The actor can type in a new name or password or choose to cancel the operation, at which point the use case ends.
6.3	**Exception Flow(s)**	User should enter correct details else he will not login
7.	**Includes**	No other use case
8.	**Preconditions**	The user Mobile should be connected to the internet and registered. Application must be installed
9.	**Post conditions**	The user has logged in and home screen will appear
10.	**Notes/Issues**	None

The user will use the login use to enter the application. The user will enter details and the information will be authenticated. If user is existed, the success message will return and access to the application will be granted. If the user is not in the system databases, error message of incorrect will be shown. The graphical representation of the login use is as shown in Figure 2-3.

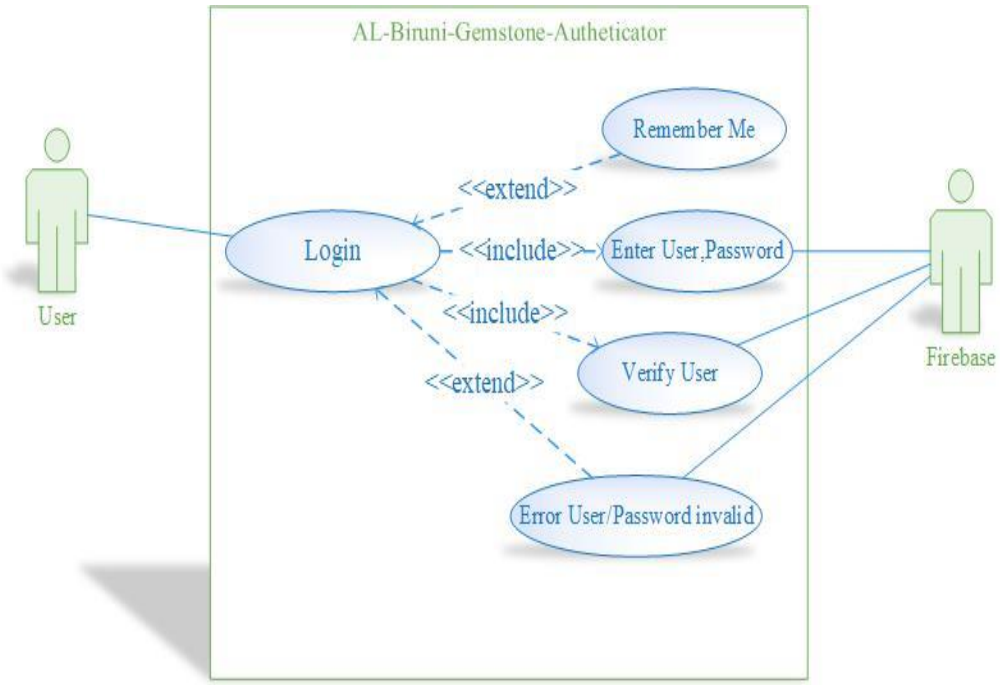


Figure 2-3 Login Use case Diagram

2.4.7.3 Forgot Password(U3)

The purpose of this use case will be used to recover password of the user.

Table 2-3. Forgot Password Use Case

	Name	**Forgot Password**
1.	**Use-Case ID**	U3
2.	**Objective**	To create a new password or recovery password
3.	**Priority**	Low
4.	**Source**	Muzamil (Developers)
5.	**Actors**	User

6.	**Flow of Events**	• Run the application. • select the Login option. • select the forgot password option • Answer Security Question • Enter New password • Press Reset password
	Basic Flow	Answer the security question and enter new password.
6.2	**Alternate Flow(s)**	No alternate flow
6.3	**Exception Flow(s)**	User should enter correct details else his account access will not be granted
7.	**Includes**	No other use case
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered
9.	**Post conditions**	The user account password has been reset and login screen will appear.
10.	**Notes/Issues**	None

The user can use the forgotten password to recover the password or update. The user will use the forgotten password enter detail and answer a security question. If the question's answer is right, then system will ask the user to enter new password. The password entered will be update if valid and success message is shown to the user screen. In case of wrong answer message will be display. The graphical representation of the system is shown as Figure 2-4.

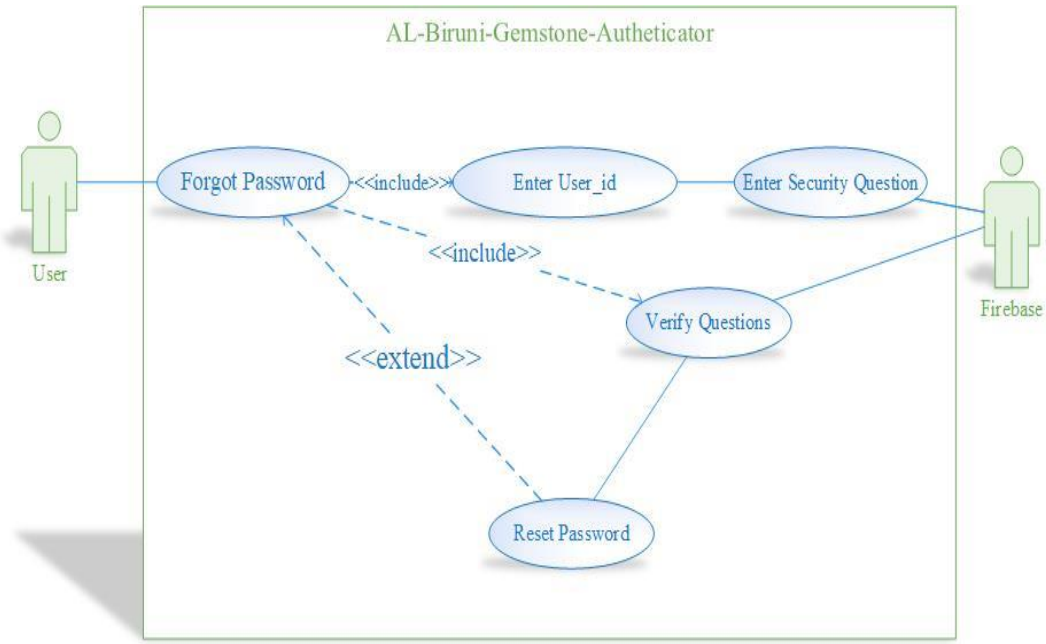


Figure 2-4 Forgotten password Use case Diagram

2.4.7.4 Reset Password(U4)

This use case will be used to reset the password for the user.

Table 2-4. Rest Password Use Case

	Name	**Reset Password**
1.	**Use-Case ID**	U4
2.	**Objective**	To create a new password
3.	**Priority**	Low
4.	**Source**	Muzamil (Developers)
5.	**Actors**	User

6.	**Flow of Events**	• Login • Go to setting • select the resetting the password • Enter old password • Enter new Password • Press the reset password button
6.1	**Basic Flow**	The information is entered, and the user password is reset and updated
6.2	**Alternate Flow(s)**	No alternate flow
6.3	**Exception Flow(s)**	No exception flow
7.	**Includes**	U2
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered
9.	**Post conditions**	The user password is reset and can be used now, login screen will appear.
10.	**Notes/Issues**	None

The user can use the reset password for the password update. The user will first login, use the reset password enter detail new password. If the password fulfills the right format for password, then system will ask the user to enter new password again. The password entered will be update if valid and success message is shown to the user screen. In case of not valid message will be display. The graphical representation of the system is shown as Figure 2-5.

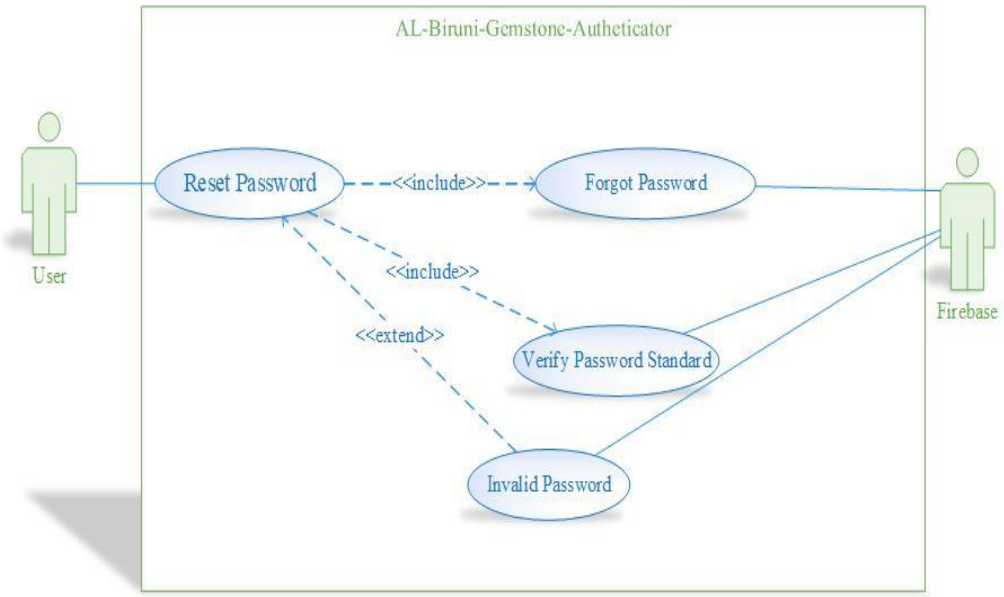


Figure 2-5 Reset Password Use Case Diagram

2.4.7.5 Capture Picture(U5)

It will be used to capture the picture for the identification.

Table 2-5. Capture Picture Use Case

	Name	**Take the Picture**
1.	**Use-Case ID**	U5
2.	**Objective**	To capture the live image of gemstones.
3.	**Priority**	Medium
4.	**Source**	Jibran (Developers)
5.	**Actors**	User
6.	**Flow of Events**	• After logging in the application • select on the capture image option • select the image or picture • Select the desired ones

6.1	**Basic Flow**	The Picture will be captured using phone camera and can be used for further purposes by the user.
6.2	**Alternate Flow(s)**	No alternate flow
6.3	**Exception Flow(s)**	No exception flow
7.	**Includes**	U2
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered
9.	**Post conditions**	The user can use live image for authentication and Identification, or authentication screen will appear
10.	**Notes/Issues**	None

Capture an image use case will help the user to take the live images of gemstone. User will first login and after successful login. Application will detect camera or check, or permission allowed for the camera. If both conditions fulfilled, user would get the access to camera and camera screen will appear. On the contradiction side error will be displayed. Upon camera access user can click images and use it in application. The graphical representation of the use case is as show as in Figure 2-6.

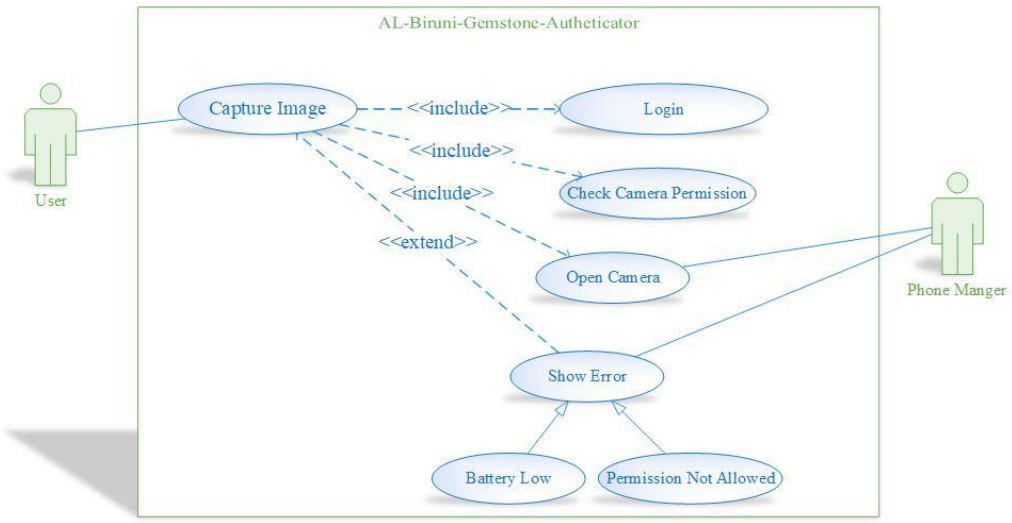


Figure 2-6 Capture Picture Use Case Diagram

2.4.7.6 Upload Image(U6)

It will be used to upload the image to the application.

Table 2-6. Upload Image Use Case

	Name	**Upload Image**
1.	**Use-Case ID**	U6
2.	**Objective**	To Upload image to the application.
3.	**Priority**	Medium
4.	**Source**	Jibran (Developers)
5.	**Actors**	User
6.	**Flow of Events**	• After logging in the application • Press on the upload image option • Select the image from the gallery • Press Upload
6.1	**Basic Flow**	The images are selected and uploaded to the application and can be used for authentication.
6.2	**Alternate Flow(s)**	No alternate flow
6.3	**Exception Flow(s)**	Images must be of support format else upload will not be done.
7.	**Includes**	U2
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered.
9.	**Post conditions**	The user can use live image for authentication and Identification, or authentication screen will appear
10.	**Notes/Issues**	None

Image use case will help the user to take the upload images of gemstone. User will first login and after successful login. Application will permission allowed for the gallery. If both conditions fulfilled, user would get the access to gallery and gallery screen will appear. On the contradiction side error will be displayed. Upon gallery access user can upload images and use it in application. The graphical representation of the use case is as show as inFigure 2-7.

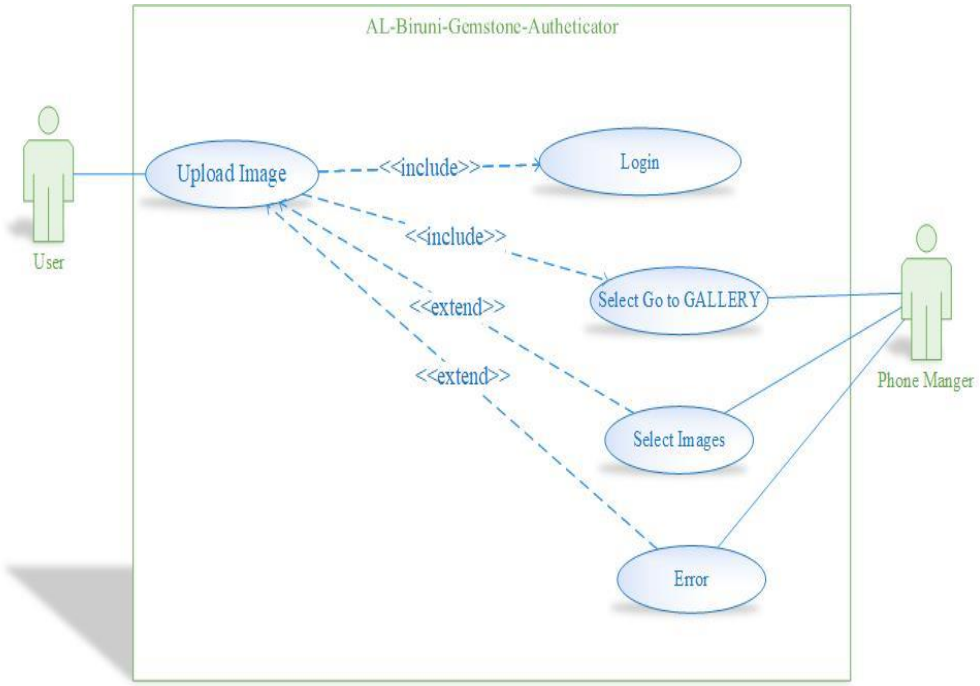


Figure 2-7 Upload Image Use Case Diagram

2.4.7.7 Edit info(U7)

It will be used to edit the information.

Table 2-7. Edit Info Use Case

	Name	**Edit Info**
1.	**Use-Case ID**	U7
2.	**Objective**	To Edit the User Information.

3.	**Priority**	Low
4.	**Source**	Jibran Ali (Developers)
5.	**Actors**	User
6.	**Flow of Events**	• After logging in to the application • Press on the setting option • Press the edit info option • Enter information • Press Edit
6.1	**Basic Flow**	Edited information will be edited. Existed information is edited, and it will be updated.
6.2	**Alternate Flow(s)**	Information entered is not valid, error will be display and user can now enter information again
6.3	**Exception Flow(s)**	Information must be valid else no editing will take place
7.	**Includes**	U2
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered
9.	**Post conditions**	The user account password has been reset and login screen will appear.
10.	**Notes/Issues**	None

The user will first login and then use this case. After successfully login user will go to the setting. Enter the new details, the details will be verified. If the information is correct, information will be updated. Else error message will be displayed to the user screen. The graphical representation of the use case is as show in Figure 2-8.

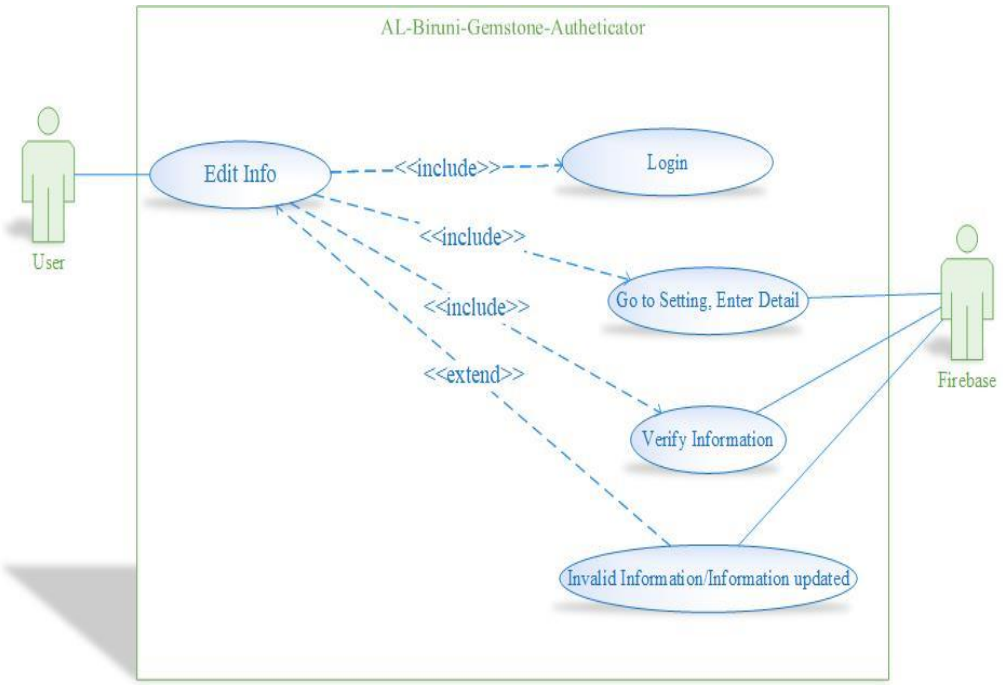


Figure 2-8 Edit information Use Case Diagram

2.4.7.8 Change Image(U8)

It will be used to change the selected image.

Table 2-8. Change Image Use Case

	Name	**Change Image**
1.	**Use-Case ID**	U8
2.	**Objective**	To Discard the selected images.
3.	**Priority**	Low
4.	**Source**	Jibran Ali (Developers)
5.	**Actors**	User
6.	**Flow of Events**	• After logging in to the application • Press on the change image option • Select images for discarding

		• Press discard
6.1	**Basic Flow**	Select the images for discarding and press discard. Existed images are discarded.
6.2	**Alternate Flow(s)**	Invalid response, error displayed, and user can use this feature again
6.3	**Exception Flow(s)**	Images must be selected for discarding else error message will display
7.	**Includes**	U2, U5, U6
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered and images are selected.
9.	**Post conditions**	Images will be discarded, and home screen will appear.
10.	**Notes/Issues**	None

Change image will be used to discard the selected images. If no images were in the selection state any error message will display. Upon success, images will be discarded, and home screen will appear.

The graphical representation of use case is as shown in Figure 2-9.

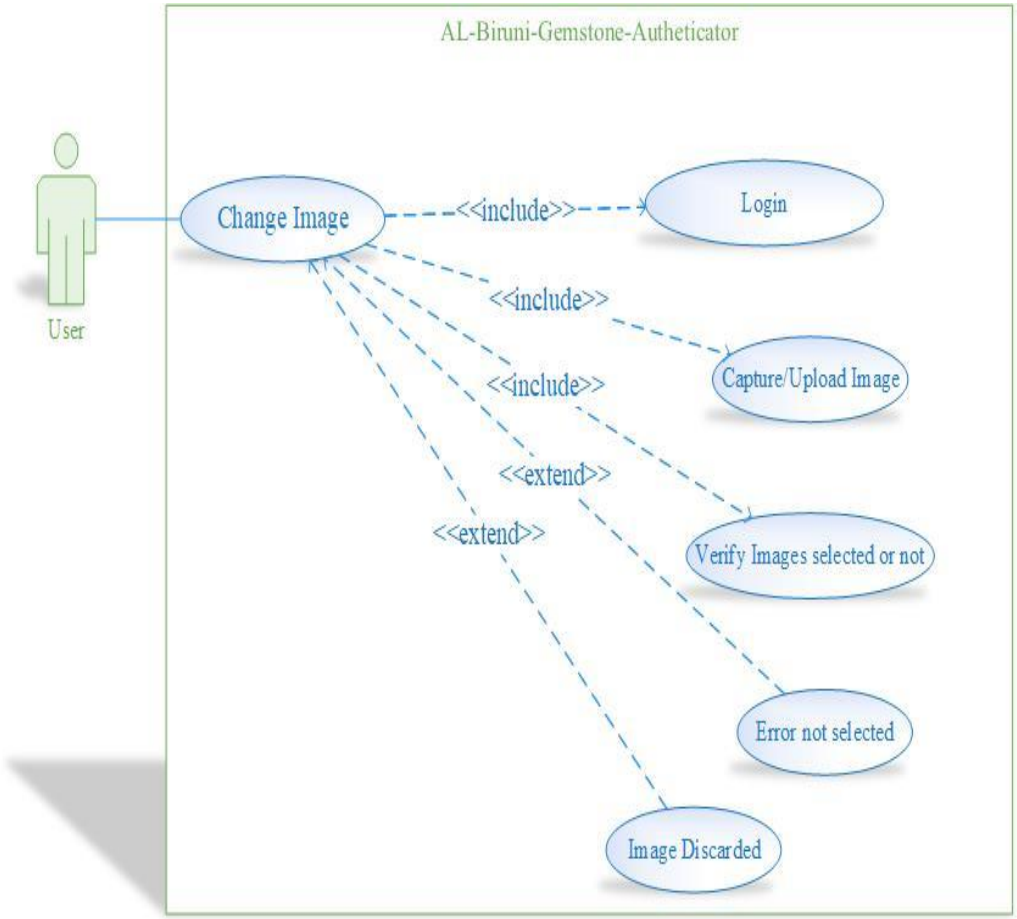


Figure 2-9 Change Image Use Case Diagram

2.4.7.9 View Identification Result(U9)

It will be used to view the identification result.

Table 2-9. View Identification Result Use Case

	Name	**View Identification Result**
1.	**Use-Case ID**	U9
2.	**Objective**	To View the results about gemstones in the image.
3.	**Priority**	High
4.	**Source**	Muzamil (Developers)

5.	**Actors**	User
6.	**Flow of Events**	• After logging in to the application • Select Images/ Press Images • Press Generate the results • Press the View Results
6.1	**Basic Flow**	Results can be viewed by pressing on view results option.
6.2	**Alternate Flow(s)**	No alternate flow
6.3	**Exception Flow(s)**	Images selected must be valid else error message will display
7.	**Includes**	U2, U5, U6
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered and images are selected
9.	**Post conditions**	Result will be generated, and result screen will appear.
10.	**Notes/Issues**	None

This use case will be used to perform the identification on the images. Below is the sequence

1. User login
2. User Select Images
3. Users ask for identification
4. If Images are selected and valid go to 6
5. Else images not selected or invalid throw an error and go to home screen
6. Identification takes place and results are generated
7. Results are showed on the user screen
8. End

The graphical representation of the use case is as shown as in figure Figure 2-10.

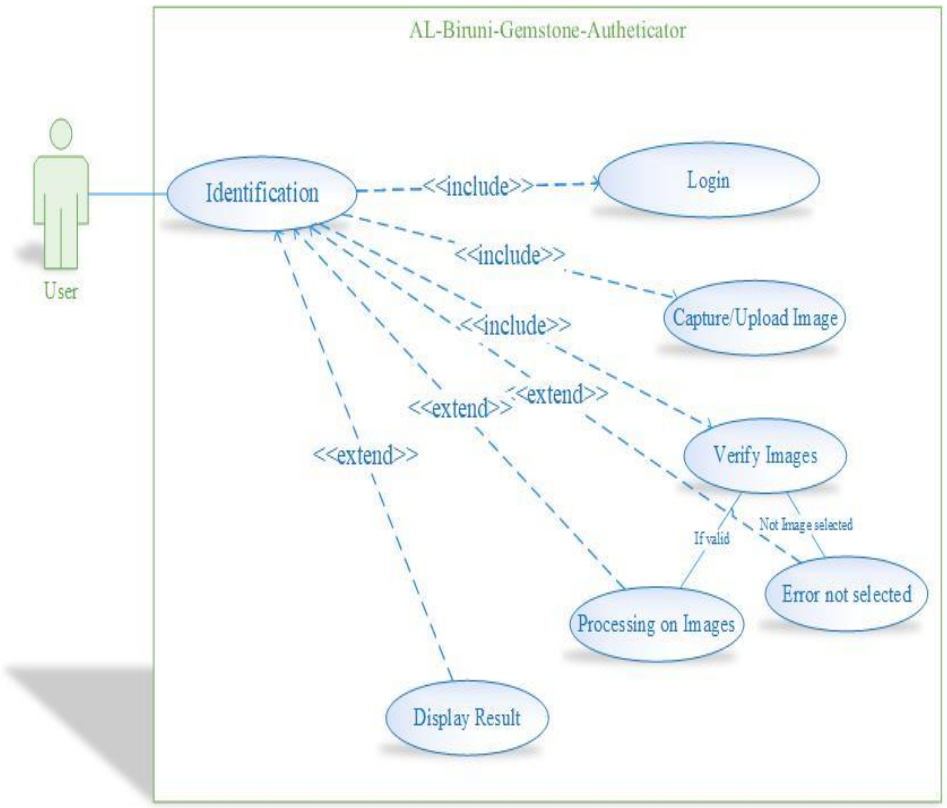


Figure 2-10 Identification Use Case Diagram

2.4.7.10 Save Results(U10)

It will be used to save the results to the external memory.

Table 2-10. Save Result Use Case

	Name	**Save Results**
1.	**Use-Case ID**	U10
2.	**Objective**	To Save results in the pdf form or image form.
3.	**Priority**	High
4.	**Source**	Jibran Ali (Developers)
5.	**Actors**	User

6.	**Flow of Events**	• After logging in to the application • Enter Detail • Select images • Press the View Results • Press Save Results • Enter Details • Press Save
6.1	**Basic Flow**	Results can be stored by pressing on the save result button and further selecting the desired format.
6.2	**Alternate Flow(s)**	If non supported format entered, error message will display, and user can use this feature again
6.3	**Exception Flow(s)**	FORMAT must be entered
7.	**Includes**	U2, U5, U6, U9
8.	**Preconditions**	The User Mobile should be connected to the internet. Application installed and running. User already registered and images are selected
9.	**Post conditions**	Results are stored in the desired format to the external memory.
10.	**Notes/Issues**	None

If the identification result is generated. Enter the desired format and press save. The result will be saved in desired format at certain location. The graphical representation of the use case is shown in Figure 2-11.

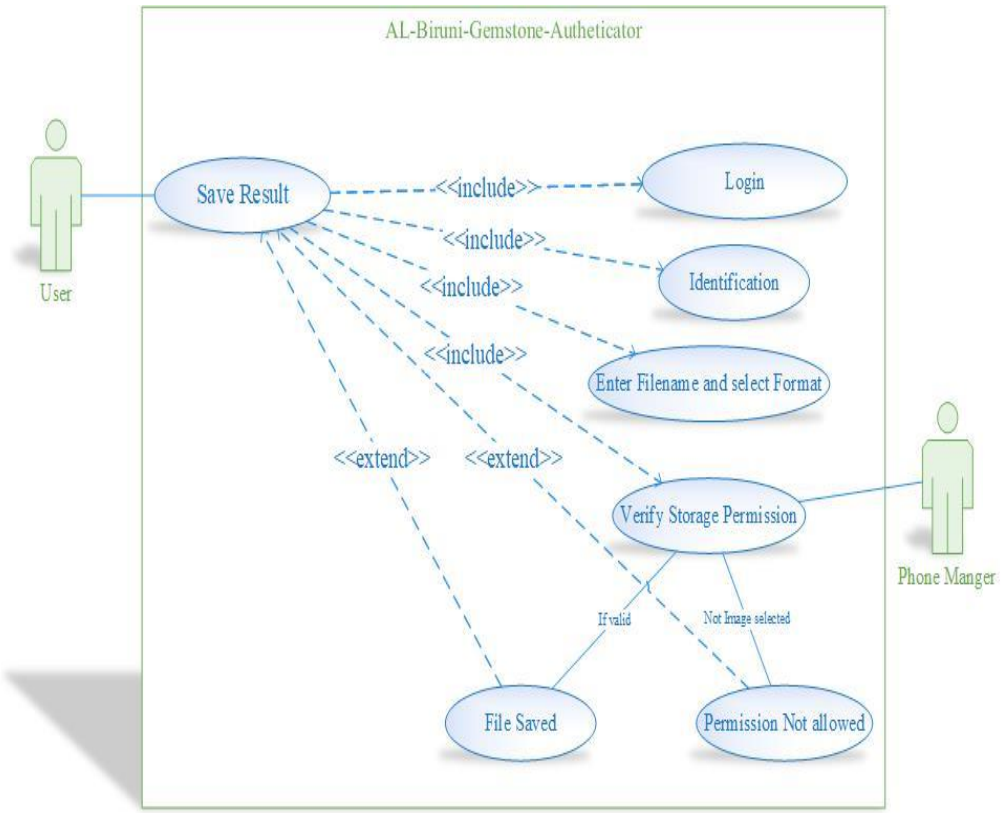


Figure 2-11 Save Results Use Case Diagram

2.5 System Features

Al-Biruni Gemstone Authentication system have two features:

- Feature 1: Identification of gemstone
- Feature 2: Authentication of gemstone

2.5.1 Identification of Gemstone

Feature 1 task is to identify the object in the image. In our feature one we are going to identify the gemstone in the images using deep learning. The identification will be done on coco dataset. The description is given below.

2.5.1.1 Description and Priority

This feature identifies the type of gemstone only two to four gemstones (Turquoise, Ruby, Agate, and Emerald) other than this four it will not identify other type of gemstone. This feature is the 2nd most important in the system. The cost of the feature will be costly calculated after its completion, till the assumption is around 1-1.2 million. The priority rating of this feature is 9.

2.5.1.2 Stimulus

Below the step-by-step sequence to use feature

1. Log in to the application
2. Select Images/ Press Images
3. Press Generate the results
4. Press the View Results

After the identification, authentication takes place.

2.5.1.3 Functional Requirements

The system shall identify the type of the gemstone. More detail regarding this feature is described in section 2.4.7.9.

2.5.2 Authentication of Gemstones:

Feature 2 will authenticate either the gemstone in the images is real or fake. This step will take place after the identification. The description is given below.

2.5.2.1 Description and Priority

This feature authenticates the type of gemstone only four gemstones (Turquoise, Ruby, Agate, and Emerald) other than this four it will not identify other type of

gemstone. Authentication is the most integral feature of our application. It cost is assumed around 1-1.5 million. Its rating is 10. Its priority is highest among all the features.

2.5.2.2 Stimulus/Response Sequences

Below the step-by-step sequence to use feature

- Open mobile app
- Select option of Authenticity of gemstone
- Click or upload image of gemstone to differentiate between gemstone from mobile camera.
- After that it will tell the name of gemstone and tell either it is real gemstone or fake gemstone
- Close the app

2.5.2.3 Functional Requirements

The system shall authenticate either the gemstone in the image is real or fake correctly. Requirement for this feature is as below.

REQ-1: Images must be selected
REQ-2: Identification must have been done
REQ-3: Images must be clear and within scope of the project

2.6 External Interface Requirements

In This section details regarding user interface, communication interface is provided. In the user interface, the activities are discussed that responsible of communication between user and application. Communication interface describes the details of protocols used for communication. These are further explained in detail in below.

2.6.1 User Interfaces

The system will include multiple user interface for several purposes. The system will have the following user interfaces.

1. Login
2. Signup
3. Home
4. Forgotten Password
5. Capture image
6. Identification of gemstone
7. Upload Image
8. Save Results
9. Change image
10. Reset Image

Details regarding these interfaces has already been discuss in 2.4.7. Below is the whole use case of the system that describe all the system features and relationship between them. System use case define that there are two type of module user module and software module. User module is entering information about him and images that he wishes to identify and authenticate. The software module will identify gemstone and authenticate and return the results back to the user. Identify also include sub processes like identify shape, color and internal patten for the authentication purposes. The graphical representation of the system Use case Diagram is as shown in Figure 2-12.

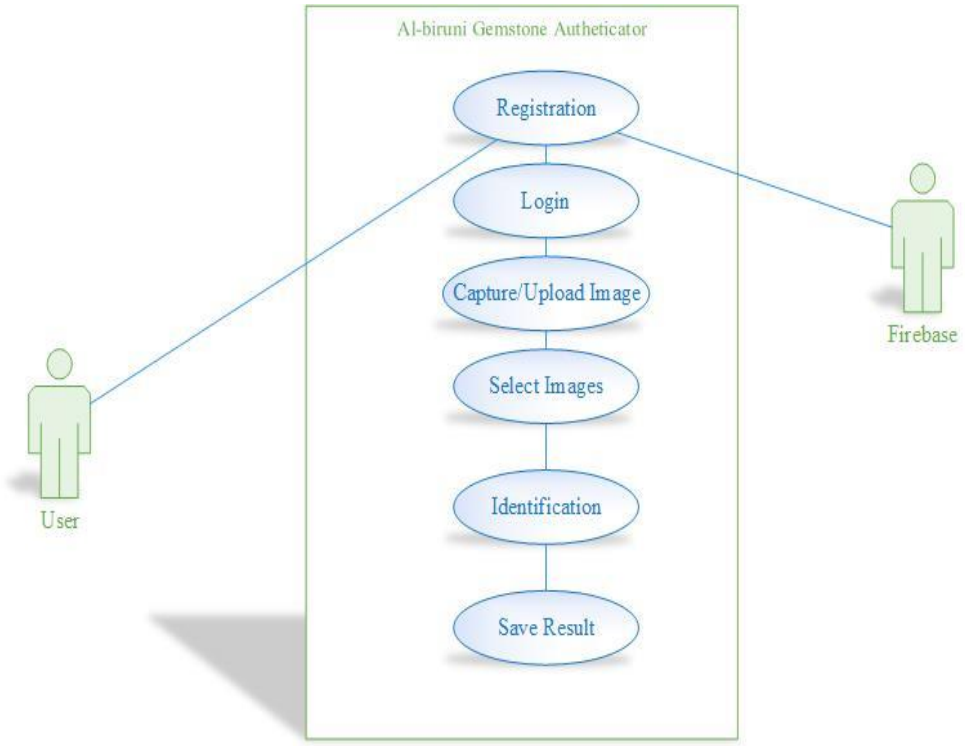


Figure 2-12.System Whole Use case

2.6.2 Software Interfaces

Following are the software used for the Al-Biruni Gemstone Authentication system application.

- Operating system: Android
- Tool: Android Studio and Co-lab
- Database: Firebase google
- Libraries: TensorFlow, Keras
- Programming Language: python and java

2.6.3 Communications Interfaces

This project supports version of android jellybean and above. The following are communication protocol use in the system.

- HTTPS for firebase Authentication
- FTP for storage transfer

2.7 Other Non-functional Requirements.

Functional requirement provides the features that system will provide. It's the nonfunctional requirement that increase the system usability. They are as important as the functional requirement. Detail of the non-functional requirement are as given below.

2.7.1 Safety Requirements

In case of any misfortunate events like hacking of developer email or database alteration. A backup copy of the database will be stored. The backup copy will be stored on the other developer email account.

2.7.2 Performance Requirements

AL-Biruni Gemstone Authenticator will run on android with jellybeans or more and good camera. The data for user information is stored in firebase hence I steady internet connection is essential.

2.7.3 Software Quality Attributes

The software quality attributes are discussed in detail in below.

2.7.3.1 Reliability

Until the user forces it to stop, or an internet connection is unavailable, the System will not fail on any grounds. According to the user's desire, any information pertaining to the identification or authentication of gemstones can be saved.

2.7.3.2 Availability

The system is available for a wide range of audience with consistency. It can be used in a wide range of environments and dynamic resources.

2.7.3.3 Maintainability

The system will generate error on unidentified actions or restricted actions performed by user inside the application. For example, upload image for identification containing out of scope objects. The database will be updated with new update of software.

2.7.3.4 Usability

The system will be designed in a such that it increases the learnability of the user. The interfaces and components used will be self-explanatory. The operability of the system shall be high. The system will follow standards that are familiar for majority

of user. The system will be designed in a such way that any ambiguous shall not remain.

2.8 Problems

In this section we will discuss what are estimated problems that we will face during the development of this project. These problems can limit the system scope or accuracy. There are two major problems in our project:

- Availability of gemstones
- Reflective Nature of gemstones

All problems are discussed below.

2.8.1.1 Availability of gemstone s

The availability of gemstones can be major problem. Every gemstone has subdivided into multiple types. It's a difficult effort to collect all the gemstone data that's part of the project's scope. There is a high possibility that not all types of gemstones will be available in the one region or whole country.

2.8.2 Reflective nature of gemstone

Many gemstones have a property that is being reflective. If the image of the shiny surface stones is taken if a reflection is shown on the surface of the stone. This can affect the purposed system accuracy. For the solution of the problem, we have use proper environment for the images. To avoid this problem, we have created some

guidelines for the dataset creation. These guidelines are defined in the dataset creation section 3.2.2.1 in chapter 3.

2.9 Literature Review

Literature is categorized under two main streams. First one is the hardware-based solution and second one is based upon machine learning algorithm. Literature review under these streams is as below.

2.9.1 Hardware

Following are the literature reviewed for the gemstone's identification using hardware.

In [2], researchers have explored the use of powder diffraction to identifying the crystalline inclusion in bulk gemstones without damaging the gemstone. They worked upon rock crystal quartz with rutile needles, rock crystal quartz with beetle legs, fake gems polycrystalline gems, moss agate, and aventurine quartz. For the identification they use laboratory X-ray powder diffraction particularly Mo Kα radiation to penetrate within compact gems technique. They achieved success for Mo radiation patterns for the identification of deeply buried inclusions and Cu radiation for the characterization of surface exposed inclusion.

In [3], researchers have explored the gemstone identification using fluorescence technique. They worked upon real ruby, fake ruby gemstones and two dye solutions (methyl orange solution and yellowish dye). For the identification they use fluorescence techniques. They achieved success.

In [4], researchers have explored the gemstone identification using handheld Raman spectrometer. They worked upon diamond emerald and sapphire stones. For the identification they compare wave numbers of the strongest Raman bands of the

minerals. They achieved fair success for strong Raman signals but not very successive for faint Raman signals.

In [5], researchers have explored the gemstone origins identification using Inclusions and spectroscopy. They worked upon emerald of all kinds (Zambia, Colombia, Afghanistan, Madagascar, Brazil, Ethiopia, China). For the identification they use Inclusion and Spectroscopy devices. They achieved below average success because of diversity categories of the emerald.

In [6], researchers have explored the gemstone identification using inclusion spectroscopy. They worked upon diamond, Diamond, Crystal, Opals, Turquoise. For the identification they use Inclusion and Spectroscopy devices. Further studies or research in underway to prove the research but proved infrared spectroscopy is one of the most promising analytical techniques.

Table 2-11. Summary of literature review related to hardware-based gemstone identification.

Ref	Research Title	Stones	Techniques Use	Results
[6] 1987	Infrared spectroscopy in gem identification.	Diamond, Crystal, Opals, Turquoise	Infrared spectroscopy	Fail
[7] 2011	Powder diffraction analysis of gemstone inclusions	single crystal (such as diamonds), polycrystalline (such as lapis lazuli), or amorphous (such as	Mo Kα radiation	Success

		amber)		
[4] 2016	The Ring Monstrance from the Loreto treasury in Prague: Handheld Raman spectrometer for identification of gemstones	Diamond, Emerald, and sapphire	Raman spectrometer	Success
[5] 2019	Geographic origin determination of emerald	Emerald	Inclusion and Spectroscopy	Fail
[8] 2020	Gemstone identification using fluorescence technique	Ruby Crystal	fluorescence technique	Success

2.9.2 Machine Learning

Following are the literature reviewed for the gemstone's identification using machine learning.

In [9], researchers have explored the development of a system that can identify the type of the gemstone. They worked upon three types of gemstones, Ruby, Sapphire, and Emerald. For the identification they use neural network working upon hue saturation value. They achieved 90.66% success.

In [10], researchers have explored classification of the gemstone on the base sharp and texture. They worked upon amber gemstone of 7 classes or kinds. For the identification they use image processing techniques. They achieved 83.5% success.

In [11], researchers have explored the gemstone identification and its feasibility using reflectance spectra coupled with artificial intelligence. They worked upon almandine, turquoise, agate, plastic, glass, dyed, impregnated, and Zachery treated. For the identification they obtained reflectance spectra of a transparent gem using an Analytical Spectral Devices spectrometer and Spectroscopy devices. Model is train upon these reflectance spectra. They achieved success.

In [12], researcher have explored the gemstones identification using its color as parameter. They worked upon ambers stones of 30 different classes. They used image texture analysis, mage classification, image color analysis, image texture analysis, machine vision for their solutions. Research aim was to sort the amber stones on the bases of its color. They achieved 88.28% success.

In [13], researcher have explored the quality of the gemstone using machine vision. They have worked upon opal stones. They used image analysis and classification techniques. Extract different characteristics from stones images. They achieved 90% success.

In [14], researcher have explored the identification of the gemstone. They have worked upon amber stones of 30 different kinds. They used classification techniques. Extract different characteristics like the mean, standard deviation, kurtosis, and skewness. They achieved 73.18 success.

Table 2-12. Summary of literature review related machine learning based gem identification.

Ref	Research Title	Stones	Dataset	Techniques Use	Accuracy
[14] 2013	Multiclass Amber Gemstones Classification with	Amber	11482 samples of 30 different	Decision tree with half half method	73.18 %

	Various Segmentation and Committee Strategies		classes of amber		
[13] 2016	Automated Opal Grading by Imaging and Statistical Learning	Opal	865 samples for each kind of Opal	Image processing, classification, and regression techniques, flash histogram	95.31 %
[9] 2017	Development of a Gemstone Type Identification System Based on HSV Space Colour Using an Artificial Neural Network Back Propagation Model	Ruby, Sapphire, and Emerald	30 samples	Artificial Neural Network Back Propagation Algorithm	90.66 %
[12] 2017	Amber Gemstones Sorting by Colour	Amber	9,008 samples of 30 different classes of amber	half, method, random forest	88.21 %
[11] 2019	Feasibility of gem identification using reflectance spectra coupled with artificial intelligence	almandine, turquoise, agate, plastic, glass, dyed, impregnated, and Zachery	12 Samples	Artificial Neural Network	93.81 %

[10] 2020	Classification of amber gemstone objects by shape	Amber	2800 samples	Image Processing, Computer vision	83.5

2.9.3 Research Gap

Both gemology and machine learning research are limited in the above works.
The stone samples utilized in the study are either identical or slightly different.
The amount of study done on this topic is quite minimal.
There hasn't been much study done on utilizing machine learning to identify gemstones. Only a few stones, such as amber and opal. have been studied in terms of gemstone identification using machine learning. There is a gap as no proper dataset is available for gemstone. There is clearly gap of gemstones used for studies and techniques used. More detail is given in below paragraph.

All the research done are of good accuracy, but dataset used is very limited. Till now Research done on gemstone are for identification or quantity check by shape, but not any research has been done on authentication. Our aim is to beat the accuracy achieved in [9]. The dataset used in this research is limited to 30 images and 3 gemstones. We are aiming create the dataset of large size and 2-4 gemstones identification and authentication

2.10 Conclusion

The field of computer vision has come a long way and have become more widely researched in recent years due to the advancements made in the field of machine learning. After 2012 deep learning has taken the uproot way. The reliability of the system highly depends upon the dataset collection and its training. We will try two or more neural network for the research purposes and accuracy dependence.

CHAPTER 3

DESIGN AND METHODOLOGY

3.1 Background

Al-Biruni Gemstones Authenticator is an Artificial intelligence-based identification and authentication system for gemstone. It uses advance technique of machine learning to make identification reliable. Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention [15].

3.2 Methodology

The proposed solution is the system of method used area of study. In our case the study area is machine learning. Methodology is lengthy process it is divided into phases. There is total 4 phases of the methodology—Literature Review and problem Statement, Benchmark dataset generation and Data pre-processing., Model Design And Model Verification& validation, Application. These Phases are sub-divided into sub-phases. The graphical representation of the Methodology is shown in Figure 3-1

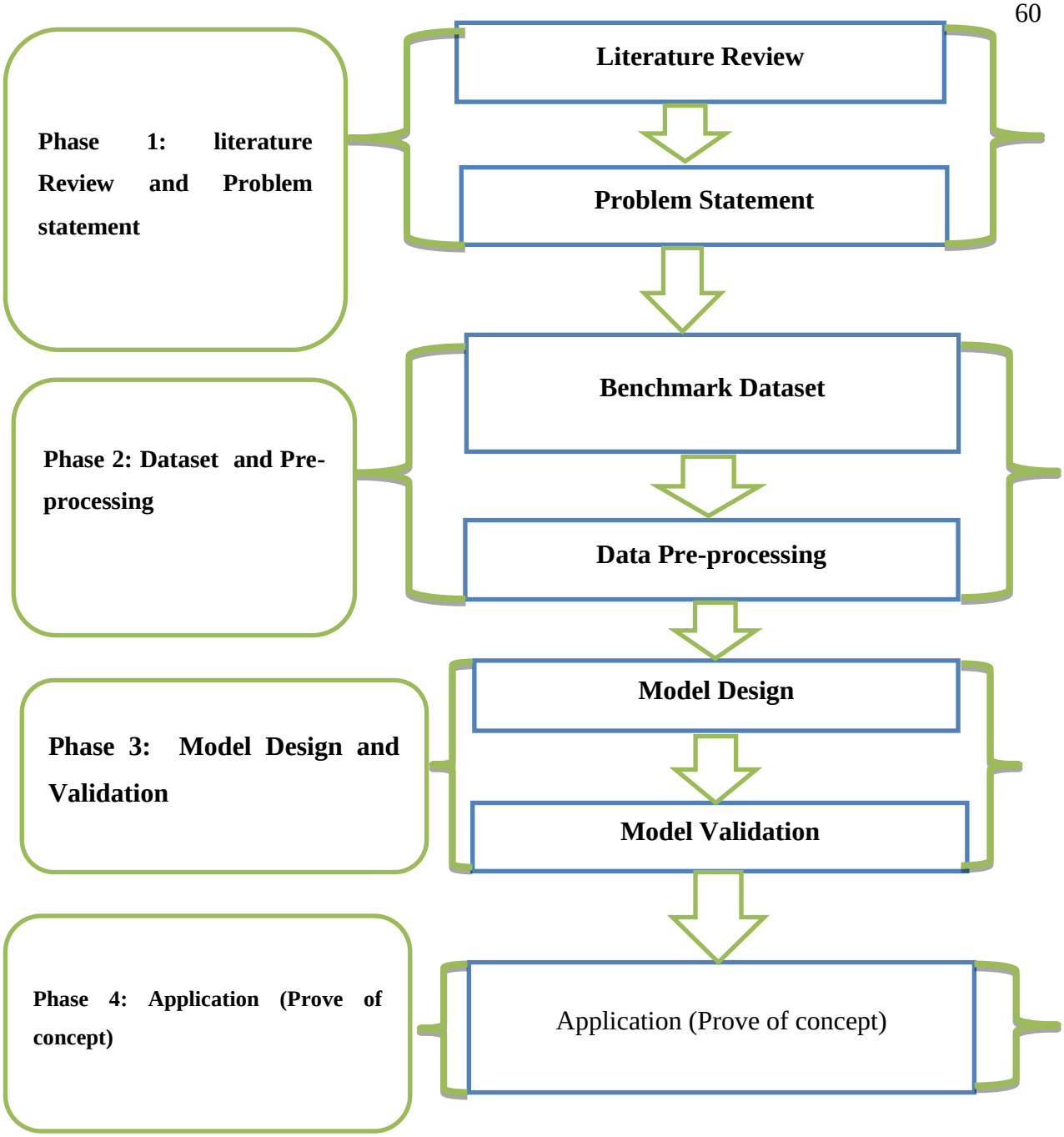


Figure 3-1. Methodology Phases

In the first Phase 1, the detail regarding literature reviewed and the problem are discussed. In phase 2, dataset generation and pre-processing operation down on the dataset are discussed. In phase 3, Model design and its validation process is explained and discussed. In the phase 4, application development which is used for proving of concept are explained. The detail of each phase is provided in the next sub sections.

3.2.1 Phase 1: Literature Review and Problem Statement

Literature review is the study of the related work already of the field. The problem statement defines the problem that audience is facing and what this solution is trying to resolve. These are first steps in solving a problem. To solve a problem, first step is to understand the problem better. The phase 1 of the methodology is sub divided into 2 subphases:

- Problem Statement
- Literature Review

The literature review is first sub phase of the Phase 1. In the sub phase, the literature review regarding gemstones identification were gathered and reviewed. The second subphases of phase 1 is problem statement. The problem statement subprocess discuss gemstone are one of most trended products round the world, product complexity is high. Because of problem high complexity, audience often get scammed. The detail of these subphases can be seen in next sub section.

3.2.1.1 Problem Statement

The proposed solution is trying to solve a common problem audience facing regarding gemstones. Since, the world of gemstones is so vast, to prevent frauds, comprehensive understanding is necessary. The product complexity is high it is ranked 838th. Gemstones was 356th most traded product in 2019. People often get scammed in this subject. The product they buy are usually fake.

To verify the gemstone identity, they need to go through the process of identification using laboratory. Gemstones identified using laboratory often loss their natural glow and Attraction. Identification fee for the gemstone is also high. The fee identification depends upon the type of gemstone. In Pakistan government sponsored laboratory charge 1000 for the identification of gemstone which is lowest in the region. The current solution is expensive and time-consuming.

Problem statement is because of the complexity of the gemstone, identification is hard. Current Solution used of this problem is hard and time consuming.

Proposed solution will authenticate between real and fake gemstones of three kinds—Emerald, Ruby, Turquoise.

The detail of problem statement is already discussed in Chapter 1.

.

3.2.1.2 Literature Review

The second subphase of the phase was to perform literature review. The literature review was done in two streams—hardware, and machine learning. There was total 14 research papers reviewed. Out of 14 research papers 6 were of hardware and 8 were related to machine learning. The techniques used in the reviewed studies are random forest, neural network, and computer visions.

To gather the literature related to gemstone, few set of search strings were used. The search string used for gathering of literature is given as below:

1. Identification of gemstones
2. Identification of gemstones using machine learning
3. Identification of gemstones using neural network

Literature gathered for review was published by different publishers. The publisher list is given as below:

- IEEE
- ICOSATS
- Applied Science Swaziland

The detailed literature review shows there is a clear research gap regarding this subject. The literature review is explained in full detail in section 2.9 Literature Review.

3.2.2 Phase 2: Benchmark Dataset and Pre-processing

The phase two of the methodology consist of two subphases. The subphase one is benchmark dataset and other one is data preprocessing. Benchmark Dataset subphase explain that the dataset was generated through two means—Visits, websites. Total number of images collected by visit were 4200 and by websites were 3000. The preprocessing section explains number of operations involved in reshaping the dataset. The operation used are Background Elimination, flash effect removal dimension reduction and cropped. The detail of each section can be seen in next subsections.

3.2.2.1 Benchmark Dataset

As machine learning project is unable to complete without data. Data is the soul behind every machine learning project. Without data machine learning project can never be successful. Since Dataset already present on the digital platform is not detailed enough for Project. There is no standard dataset available for this subject or study. The dataset available on kaggle isn't detailed enough. The images per class are under 100. The other limitation of that dataset is there are no data related to fake gemstones. These are reasons why created our own dataset. Dataset will be of multiclass type. Our dataset is primary. Primary data is the category of the data which is created by researcher him/herself. To complete the proposed solution dataset generation was essential.

3.2.2.1.1 Dataset Generation Process

As mentioned early in the section 3.2.2.1, data is of primary category. Dataset must be generated for the project. To generate the dataset, first step was to create an environment for image capturing. The purpose of the environment was to provide consistency in the dataset. The detail regarding environment is discussed in next subsection.

Controlled Environment Box

A handmade environment was created. The environment was box, inner side covered by A4 paper and small hole at the top of the box. The environment box is shown in **Figure 3-2 Controlled Environment Box**.

Figure 3-2 Controlled Environment Box

The purpose of using the environment box was to provide dataset consistency in the dataset. The goal was to provided dataset images with same background and environment. The other factors like images under same light and controlled environment. This will help the model to learn main pattern in the dataset

Without the use of the environment, the images would have different backgrounds. The other problems were like images having reflection on the surface of the gemstone. This happens because of gemstone nature and lights in the jeweller shops.

These problems can lead to the model assuming that these attributes are also part of the gemstones. Below is the image taken without environment shown in Figure 3-3.

Figure 3-3 Turquoise Image without Controlled Environment Box

As observed from the image, the image contains shadow of the gemstone. The second problem with this image is gemstone have shadow of surrounding on its surface. These two problems can affect model accuracy and makes the model learn wrong patterns.

To overcome these two above problems, environment box was created. The same type of gemstone image is captured but this time using environment. The sample is shown in Figure 3-4.

Figure 3-4 Turquoise fake image captured using controlled environment box

As observed from the above figure, the two-problem found in the image taken without environment is resolved here. Environment do help us in generating a standard dataset. To create a standard dataset, there were few guidelines to keep consistency factor in the dataset. These guidelines used for the generation of the dataset are explained in the next subsection.

3.2.2.1.2 Guidelines

To create a proper dataset, we needed consistency around the images. For this purpose, A handmade environment is being used. The dataset is created manually, and other resources are also used to add existing relevant data to the dataset. To maintain a standard for the dataset generation, few list of guidelines has been created. Detail of the guidelines is given as below:

1. Gemstones will be in raw material state
2. Gemstones front face will be used for dataset
3. Background of the dataset is white. Blank papers/Clothe is used for that purpose
4. Distance from the gemstone will be 12 inches.

5. 5-10 stones images will be used for one type of Gemstones as per dataset collection
6. Image from different angle will be taken for all the gemstones
7. Gemstone's image will be clicked in natural environment to avoid lighting factor on the shining Gemstones and to capture its natural color
8. Some Gemstone's image will be clicked under Artificial or Sunlight to capture natural random pattern inside the stones
9. Some Gemstone's image will be clicked over the jeweler's lens. This is done to get deep nature of gemstones and capture common deep factors

The dataset generation was done in two subphases—market visits, websites. The detail of these subphases is provided in the next subsection.

3.2.2.1.3 Dataset Generation through Market Visits

To generate the dataset, visiting gemstone market was an integral part of the dataset generation. As the gemstone included in the scope of the project were quite expensive and to cover different kinds and shapes of the same stone's market visit were compulsory. This was arguably the most difficult part of the study, as no jeweller and shop keeper were willing to allow us to capture gemstones images one by one. Even on requesting and offered them token money, very few of them allowed us to capture images. A moment of visit is shown in figure 3-5.

.

(a) (b) (c)

Figure 3-5 A moment with Gemstone seller during Market visits

Shop keeper or seller easily get annoyed after 5-10 minutes. As they claim that their sales were getting affected by our presence. Hardly 100-300 images were generated from each shop. Majority of the dataset created is from the shops of jeweller of our references. The variety of the stones were very limited. The expensive gemstones like emerald and ruby were hard to found and their one kind of shape or cutting were available at most. We visited 5 places for dataset generation. These 5 places were:

1. Nageena Market
2. Usman Jewellers Township Market
3. Basand Jewellers (Online Seller) Township Market
4. Kundan Jeweller International Market
5. Zeeshan Gems BOR Market

The sample generated from each place is shown in Figure 3-6.

Figure 3-6 Sample images captured during market visits.

Turquoise stone was available in majority because of its high demand and average price. All the shops have both variety of turquoise real and fake. Because of the shine issue and limited time, we were able to get only handful of good images. To increase the depth of the dataset, purchased few fake gemstones to explore all the possible angle of the gemstone. Images dimensions were different because different cameras and stone angles were used during dataset generation. Majority of images have dimensions of 4000 x 3000 and 6944 x 9248 or vice versa. Around 4315 images were collected during these visits.

The number of images generated during market visit were 4315. The classes of dataset whom images were generated are:

- Turquoise
- Fake Ruby
- Fake Emerald
- Fake Turquoise

The images generated for fake turquoise is 2600, for turquoise 600, for fake ruby 1023, Fake Emerald is 1092. The image distribution is shown in **Table 3-1. DATA generated through market visits image distribution**.

Table 3-1. DATA generated through market visits image distribution

Gemstone	Fake	Real	Total
Emerald	1092	0	1092
Turquoise	2600	600	3200
Ruby	1023	0	1023
Total	3715	600	4315

3.2.2.1.4 Dataset Generation through Websites

Visits do help us in creating efficient dataset. But expensive gemstones like emerald, ruby variety were very limited only one kind of cutting or shape was available for

these expensive gemstones. To avoid this limitation of the dataset, we used online website to collect the data. [16] is used to collect emerald gemstone. Images are generated from videos taken from this source. Then Videos were run on VLC media player to take screenshot per frames every 0.5 seconds. Similar procedure was done using [17]. Ruby real and turquoise 40% images are generated using this source. The dimension of these images is 1920 x 1088. The sample image of the dataset created through website is shown in Figure 3-7.

Figure 3-7 Image generated from video collected from [19]

The number of images generated during market visit were 4315. The classes of dataset whom images were generated are:

- Turquoise
- Emerald
- Ruby

The images generated for turquoise 490, for ruby 1007, Emerald is 1003. The image distribution is shown in Table 3-2.

Table 3-2 Dataset generated through website image distribution

Gemstone	Fake	Real	Total
Emerald	0	1003	1003
Turquoise	0	490	490
Ruby	0	1007	1007
Total	0	1500	1500

3.2.2.1.5 Character statics of Dataset

The benchmark dataset contains 6000 images. The dataset is consisted of 6 classes. The format of the images in the dataset is JPG. The dataset is completely balanced. The dimension of the dataset is 400 x 400. The rules used to create the dataset are provided in section above 3.2.2.1.2. The classes of the dataset are as below:

1. Emerald
2. Fake Emerald
3. Ruby
4. Fake Ruby
5. Turquoise
6. Fake Turquoise

The dimension of the dataset is 400, 400,3. The images colour format is RGB. The dataset is generated from different cameras usually more than 50 mega pixel cameras. The dataset is generated from the jeweller shops around the Lahore. Some part of the dataset is generated from frames of videos download from websites. The detail regarding dataset format is shown in 3.2.2.1.5.

Table 3-3.Characteristics of the dataset

Dataset Category	Multiclass

Number of Classes	6
Dimension of the dataset	400,400,3
Format	JPG
Number of Images	6000
Images per class	1000
Colour	RGB
Dimension of Images	2D
Dataset Size	236MB
Background Colour	D8c9c9

Initially images in dataset were 7200. Images less efficient were discard and only optimal were selected for the dataset. There are Total 6000 images in the dataset. The images per class is 2000, 1000 for fake and 1000 for real. The dataset is completely balanced. The detail of the image distribution is shown in

Table 3-4. Dataset Images distributions

Gemstone	Fake	Real	Total
Emerald	1000	1000	2000
Turquoise	1000	1000	2000
Ruby	1000	1000	2000
Total	3000	3000	6000

The emerald samples contain 1000 images of real emerald and 1000 images of fake emerald. The same distribution of images is for ruby and turquoise. The histogram of the dataset image distribution is shown below in Figure 3-8.

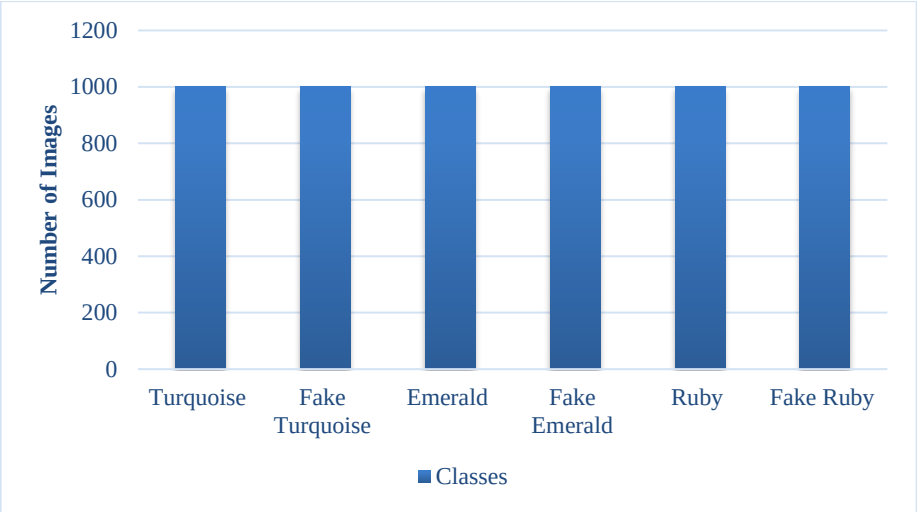


Figure 3-8 Dataset Classes Histogram

The dataset images were reshaped into 400x400. Since images were generated from different camera and different angles. There were 3 dimensions of the images—9248x6944,6944x9248, 4000x3000 ,3000x4000. The detail of reshaping process is defined in detail in section 3.2.2.2.

The dataset generation is divided into two streams—Through visit, through websites. In the first stream, dataset is generated by visiting jeweller and gemstone seller markets. To increase the depth of the dataset, in the second stream dataset is generated by using digital gemstone ecommerce platform. Videos are download, via frames images are generated and added to the dataset. Both the streams are explained in detail in the next subsections

3.2.2.2 Data Preprocessing

The second section of the phase 2 is data preprocessing. Dataset is generated primary and some of it is taken by frames using video downloaded from [18] and [19]. To cover up different angle of gemstones and use of different phone cameras, different dimension of images was generated. There was a difference of background and dimensions. Many Stones also have reflection of the light on their surface. This

problem can cause our model to make model inaccuracy. The tool used for image pre-processing is adobe photoshop. The pre-processing is done using 4 operations. These operations are given below:

1. Background Elimination
2. Flash Effect
3. Dimension Reduction
4. Cropped

Detail of these operation is provided in next subsection.

3.2.2.2.1 Background Elimination

The background for the dataset is white. The background for the dataset generated through visits contain white background for all images. But this statement is not true for dataset through websites. The videos were recorded at different environments, with different background and different lights. The background elimination was necessary to make the dataset consistency. The background decided for the dataset is d8c9c9. A colour layer of value d8c9c9 is used. The tool used for pre-processing is adobe photoshop. Below is sequence of the operation performed for background elimination:

1. Create a layer copy of the image
2. Use adobe select subject feature
3. Create fill layer
4. Create a solid colour layer of value d8c9c9
5. Move fill layer to coloured layer
6. Save the image as copy. In jpg

The adobe offers reusability if the operation to be performed are similar. Images needing background elimination required similar operations as above. Action was record for one image and then run on the all the images needing

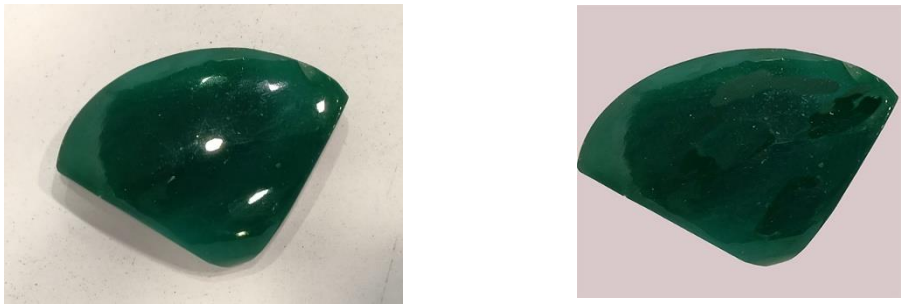
background elimination. After all the images goes through background elimination, to save all at once image processor was used. The image quality selected was 12 which highest in terms of quality. Background elimination operation is shown in Figure 3-9.

(a) (b)

Figure 3-9 Background of images are eliminated (a) before elimination (b) after elimination

3.2.2.2.2 Flash Effect Remove

The presence of extra light in jewellery shops create a reflection on the surface of the gemstone. Few of the gemstones avoid this problem because of the environment and their low reflective nature. Few gemstones like emerald which have high reflective property still gets reflection on their surface. This can make the model inaccurate. To avoid this, flash effect remove operation was performed on these kinds of images. Adobe photoshop was used for this purpose. Pattern stamp feature was used for the flash effect remove operation. This enables to repeat pattern found on the surface of the stone onto the other part or surface of the stone. The images before (A) and after (B) the flash effect remove operation are shown in Figure 3-10.

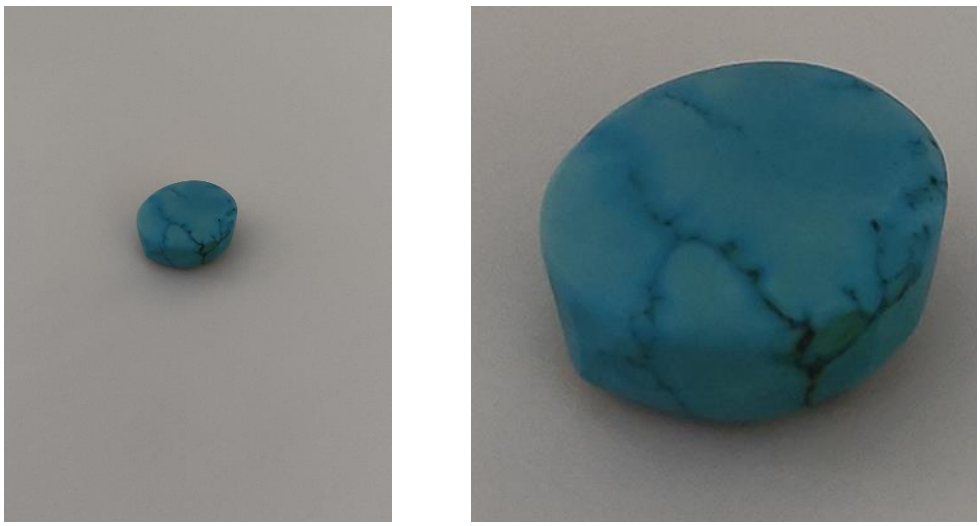


(a) (b)

Figure 3-10 Reflection of light eliminated (a) before elimination (b) after elimination

3.2.2.2.3 Cropped

The images generated were from different angles to increase the number of samples in the dataset. Because of this many samples in the dataset were containing extra objects and space other than gemstone. To reduce image size and easiness for model images were cropped. The cropped operation was performed to extra pixels in the image. Remove those pixels which are unnecessary. The image before (a) and after (b) this operation is shown in figure 3.11.



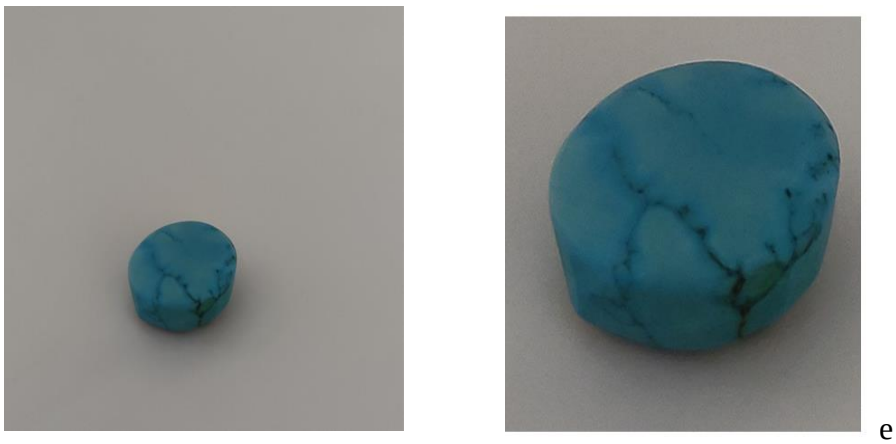
(a) (b)

Figure 3-11. Extra area of images changed (a) original image (b) cropped image

3.2.2.2.4 Dimension Reduction

The initial dataset contains images from different angles and different cameras because of this dataset images dimensions were different. Their size was also large. The average size of the image was 8 megabytes. Below feeding to the network images must be of same dimensions and small for fast processing. The images dimension was 9244x6948 and vice versa. Working on these large dimension makes the network slow. To avoid this images dimension was reduced to 400x 400 using adobe. This step involved use of crop tool of the adobe. The size for cropping was 400x 400. Dimension reduction was done using cropped operation. Adobe crop feature allow us to enter 400x400 and 92 pixel per inches. The sample before and after (b) dimension reduction is shown in (a) (b)

Figure 3-12.



e

(a) (b)

Figure 3-12. Dimension of images changed (a) original dimension (b) resized dimension

3.2.3 Phase-3: Model design and Validation Verification

In this phase 3 there are two section, Model Design and Mode verification & Validation. Model defines of the model. Total number of layers for model design are 12. The technique used for model design is convolution neural network. Model validation methods are cross validation and testing. The detail of these subphases is provided the next subsection:

3.2.3.1 Proposed model design

Technique used for model design is deep convolutional neural network. The network model consists of 12 layers -------- 4 convolution layers, 4 max pooling layer, 2 dense layer, 1 flatten layer and 1 dropout layer to avoid overfitting. The input shape for the network model is (400,400,3). 400 is length and width and 3 is image type. The colour format of the images is RGB. The network is implemented using python programming language version 3.7. The library for the implementation is Keras. The neural network is Convolution neural network. The CNN is build using keras Sequential model. There are two ways to build models in keras, functional and sequential. Sequential API allow you to build model layer by layer. But there are few limitations like does not allow you to create a model that share layers or that have different inputs and outputs. On the other hand, functional API allows you to create a model with lot of flexibility. Functional API of keras offers you define a model where layers are connected to more than previous and next layers. The architecture of the model is Figure 3-13.

Model: "Sequential"

Layer (Type)	Output Shape	Param #
Conv2d (CONV2D)	(None, 398, 398, 32)	896
Max_pooling2d (MaxPooling2D)	(None, 199, 199, 32)	0
Conv2d_1 (CONV2d)	(None, 197, 197, 64)	18496
Max_pooling1d (MaxPooling2D)	(None, 98, 98, 64)	0
Conv2d_1 (CONV2d)	(None, 96, 96, 128)	73856
Max_pooling2d_2 (MaxPooling2D)	(None, 48, 48, 128)	0
Conv2d_3 (CONV2d)	(None, 46, 46, 128)	147584
Max_pooling2d_3 (MaxPooling2D)	(None, 23, 23, 128)	0
Flatten (Flatten)	(None, 67712)	0
dropout (Dropout)	(None, 67712)	0

Figure 3-13 Proposed Model Architecture

The architecture of the model explained layer by layer in the sub section.

3.2.3.1.1 Layers

Layer defines the topology of the model. This topology of the network can define hypothesis space. There are 4 types of the layers involved in the building model architecture.

1. Convolution Layer
2. Max Pooling
3. Flatten
4. Dense
5. Dropout

The first layer of the model is Conv2D layer. This layer creates a convolution kernel that is convolved with layer input to produce tensors of output. This layer takes few

parameters like kernel size, number of filters, activation function and input shape if Conv2D is first layer of the model. In our case, number of filters is 32 as it is recommended to use lower number of filters at the start and higher number of filters later. The activation function is rectified logical unit which is considered best for start and hidden layers. The kernel size is (3, 3) which must be odd. At the last input shape is provided (400, 400, 3).

There is total 4 Conv2D layers in our model. The 1,3,5,7 layers are Conv2D as shown above in Figure 3-13. The second layers have number of filters 64, kernel size of (3, 3). Conv2D layers in between will have more filters than the early Conv2D layers but fewer filters than the layers closer to the output. The 3 and 4 Conv2D layers has same kernel size as previous Conv2D layers but higher number of filters. The activation function is same for all the Conv2D layers.

The second type of layer used in model design is Max pooling. The role of max pooling: to aggressively down sample feature maps, much like strided convolutions. It is conceptually like the convolution layer, but convolution is done on local patches using linear transformation while in max pooling samples are down sampled via hardcoded max tensor operation. Their purpose is to keep the max value that can makes identification or authentication easier for the model It keeps the most useful feature of the image and discard the rest. The max pooling is done on the window size of (2,2). The 2,4,6,8 layers in the model are max pooling. Their usually used after convolution to down sample the learned output tensor.

Convolution layers may be effective while working for visual analysing, but in terms of classification we cannot use them. To use classifier, we must use dense layer on the stack of Convolution layers. There is small difference between their working mechanisms, convolution working on 3-dimensions tensors while dense layer work on 1-D vectors. To convert multiple dimension tensor to 1-Dimension we use flatten layer.

After converting 3-d output tensor to 1-D, dense layers are added on to the top of previous all layers. The purpose of using dense layers is to use classifier with convolution layer. The activation function used for the last layer is soft max for

multiclassification. The number of neurons in the last layer must be equal to the number of classes used in the dataset. The number of classes are 6. The last layer neuron is 6.

As shown in Figure 3-13, there is use of drop layer before dense layers. The intention behind using this layer is to perform regularization on network. It randomly drops out output values. The values are set to zero. The number of dropping out is equal to the value provided to dropout layer. 0.5 value is used for dropping out for our model. 50% of the output tensor would be randomly set to zero.

There are total of 12 layers. There is total 4 convolution layers out of 4. The 1,3,5,7 layers are convolution. There 4 max pooling layers out of 12. The layer number 2,4,6,8 are the max pooling layers. There are 2 dense layers out of 12. The layer number 11 and 12 are dense layers. There is 1 drop layer. The 9^{th} layer is drop out layer. The layer number 10 is flatten layer. The graphical representation of the model layers order is shown in Figure 3-14 Layers Graphical Representation of model.

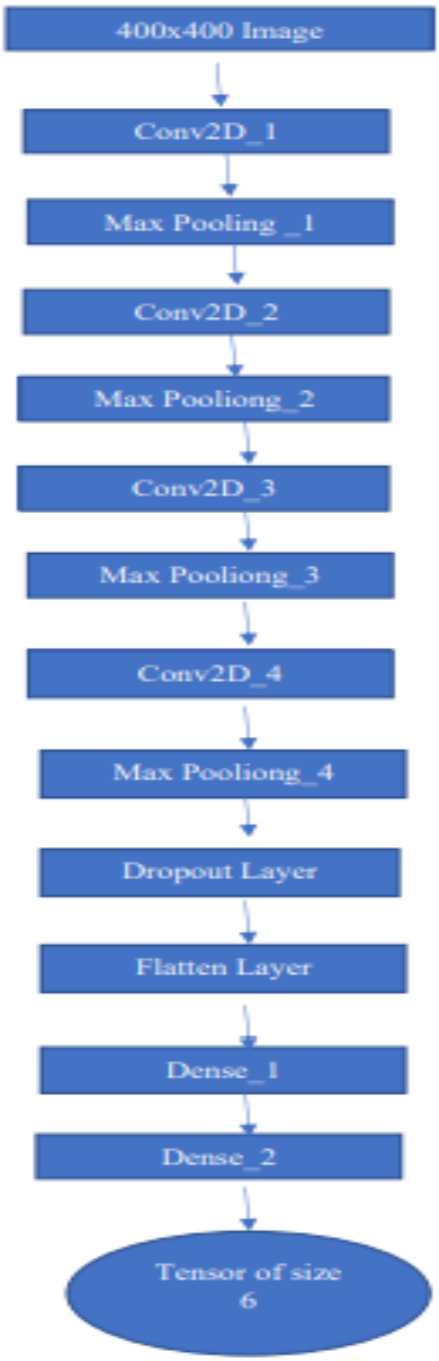


Figure 3-14 Layers Graphical Representation of model

3.2.3.1.2 Hyper Parameters

. There are few parameters that made model a success or failure. The parameter that can affect model is as given below:

- Epoch
- Batch Size

- Optimizer
- Learning Rate
- Loss Function
- Activation Function

All these parameters are explained in the next subsections.

3.2.3.1.3 Optimizers

Optimizers in machine learning are used to tune the parameters of neural network. The main purpose is to minimize the cost function. There are many optimizers available to use. The choice of the optimizer is not an easy choice, but there are two main categories. This line of boundary is drawn on the bases of their functioning. There are two types of optimizers, gradient descendant, and adaptive optimizers. In the gradient descendant, the learning rate is needed to tune manually while in adaptive it's tune automatically. In this study, the optimizer used is ADAM. This is considered as the best optimizer.

3.2.3.1.4 Learning Rate

The learning rate is a hyperparameter that rules how much to switch the model in reply to the estimated error each time the model weights are updated. Choosing the learning rate is challenging as a value too small may result in a long training process that could get stuck, whereas a value too large may result in learning a sub-optimal set of weights too fast or an unstable training process. The learning rate may be the most important hyperparameter when configuring your neural network. Therefore, it is vital to know how to investigate the effects of the learning rate on model performance and to build an intuition about the dynamics of the learning rate on model behaviour. It

comes down to experimentation and instinct. The learning rate for our model is 0.0001, 0.001, 0.00025

3.2.3.1.5 Batch Size

Feeding complete dataset at once is tough nut to crack. It's expensive and inefficient. The breakdown of data into small and equal parts is the batch size. The dataset is divided into equal parts and then sent forward to neural network. The batch size for the model must be in even. The lower batch size lower model speed. The common batch sizes are 2, 4,8,32. The batch size for our model is 8.

3.2.3.1.6 Epoch

The number of epochs can be defined as the number of times training data will be run on the model during training phase. The purpose of this to makes the model better learn by working on its multiple number of times. There is no standard where model will give it high performance and where low. It is advised to use large number of epochs to cover all the possibilities. The standard number of epochs are 200. The number of epochs used for our model is 150. For the experiments the number of epochs is 50.

The model design and its architecture are explained in detail in section Proposed model design 3.2.3.1

3.2.3.1.7 Activation Function

An activation function is a function that is added into an artificial neural network to help the network learn complex patterns in the data. The Activation function used were softmax and relu.

The softmax function outputs a vector of values that sum to 1.0 that can be interpreted as probabilities of class membership. It is related to the argmax that outputs a 0 for all options and 1 for the chosen option. Softmax is a lighter version of argmax that allows a probability-like output of a winner-take-all function. As such, the input to the function is a vector of real values and the output is a vector of the same length with values that sum to 1.0 like probabilities.

Relu for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance.

3.2.3.1.8 Loss Function

Classification problems involve predicting a discrete class output. It involves dividing the dataset into different and unique classes based on different parameters so that a new and unseen record can be put into one of the classes. The loss function used for our problem for classification Categorical cross entropy loss.

Categorical Cross Entropy loss is essentially Binary Cross Entropy Loss expanded to multiple classes. One requirement when categorical cross entropy loss function is used is that the labels should be one hot encoded. This way, only one element will be non-zero as other elements in the vector would be multiplied by zero. This property is extended to an activation function called softmax.

3.2.3.2 Validation and Verification

Our project is to identify and authenticate between real and fake gemstones, so we need pictures of gemstones to train our machine. We used Convolutional Neutral Network. A Convolutional Neural Network is a Deep Learning algorithm which can take an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. It has Inner and outer function activation function. The model sees and learns from the data. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. The validation of the project is done using two techniques. One is cross validation and other is testing data. Both are explained briefly below.

3.2.3.2.1 Cross Validation

Cross validation is used when dataset is divided into K-groups. One of them is used for training and other for the testing. The main speciality of the K-fold is it divide the dataset K-splits, during each iteration a new n^{th} split of K is used. The K-split of the cross validation for this study is 5. Cross validation helps to prevent model to go under fit or over fit. The main cause of the overfitting is model rather than learning the pattern instead remember training data. This in turn cause the model to perform low on test or validation data than training data. In Cross validation, each time a new set of testing and training data is used. This mechanism makes the model to avoid or prevent overfitting and under fitting which in turn increase accuracy.

3.2.3.2.2 Testing

After the completion of the training life cycle model, testing phase begin. This portion of the data is to evaluate the new trained machine learning model. Model performance is better checked on this portion as this set of datasets is unseen for the model.

3.2.4 Phase 4: Application (Prove of concept)

Previous phases of the methodology were focusing on problem and tool technique used for model creation and validation. Model do solve the problem but to be used in daily life. Model must in easily accessible form like android application. The phase 4 focus on implementation of application. The application is designed only as prove of concept. The detail of this subphase is provided in next subsection.

3.2.4.1 Application

Model of the machine learning needed to be integrated with application or must be deployed on websites for access. Our model is integrated with android application. The android application is developed using android studio. The programming language used for the implementation is Java. The application is developed only as prove of concept. The application proves two main functionalities, identification, and authentication of the gemstone. The stones included in the scope of the project will get verified. The complete detail of application feature and requirement is already discussed in CHAPTER 2.

App will take input in the form text and images. The result of every prediction will be saved in the application. Use cases of application are already explained in chapter under section 2.4.7

3.2.4.2 Workflow of Application

Gemstones are identified and authenticated by system when it receives an image from a camera or gallery, passes that image to model that has been trained to perform authentication. Results are generated and saved in the application. The figure 3-16 describes the overall workflow of gemo system.

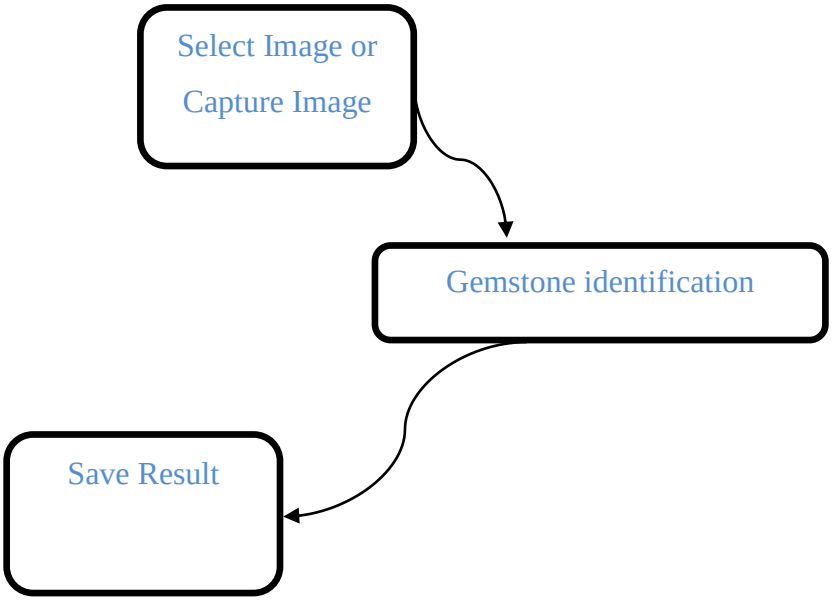


Figure 3-15 Workflow of gemstone identification (gemo application).

There are three modules involved in the work follow of the system. These modules are

- Select or capture Image
- Gemstone identification model
- Save Result

These modules are explained in detail in the next sub sections.

3.2.4.2.1 Select / Capture Image

The user can use this module to enter his input image to the model and generate its result according. The input type is image as model is working on visual data. The

image can be selected from gallery or taken using camera. As model is simple and input shape is 400x400. The image must be in 400x 400. Before sending the image to the model, image is reshaped in to 400x400. The selected images can be discarded and new can be selected. There are two options—Pick image, take image. The pick image will open gallery and gives the user choice to select one image. The take image would open camera if permission allowed. Else first permission will be asked and then camera will be able to take and send back that image to application.

3.2.4.2.2 Gemstone identification model

This module is used to send the selected images to the model. This module includes select or capture image as its sub module. The selected images from the user will be send to model if user press the submit button. The image in the form bitmap is send to tensorflow lite model. The model will send its result or output tensor according to the image results. This module is the main module of the application. The user must be login first to use this feature. The model used in this module is of tensorflow lite format.

3.2.4.2.3 Result generation

In this module, the output tensor or result send by the model is translated and saved in the application. The results are identified by the identifier added at the module 1. The result return is the type of gemstone. The result can be view inside the application under result tab. This is last module in the workflow of the application. The results can be saved from inside the application to the external storage of the phone.

3.2.4.3 System interface on Android

The interface would receive Image from camera or gallery and then identify for the type of gemstone present in those images. Results are automatically saved inside the application. The list of actions to be performed on end system are listed below:

- Initiate System
- Close System
- Edit Information
- Take Image
- Select Image
- Authentication and identification
- View Reports

3.2.4.4 Success Criteria

The following are the criteria of success for the proposed system:

- Correct Identification of type of gemstone
- Authenticate between real and fake gemstone

3.2.4.5 Sequence Diagram

The Sequence diagram of the project are list below.

- Login
- Sign Up
- Forgot Password
- Reset Password
- Take Picture

- Upload Image
- Edit Info
- Change Image
- View the Identification Result
- Save the Results

They described in detail below.

3.2.4.5.1 Signup

If a new User is to be registered into the system, they must enter the signup details on the signup activity and press signup button. The system would validate the details entered, then sends the data to the server side. Further user created notification is generated and shown to the User, if wrong details are entered, the system would then generate an error message as shown in Figure 3-16.

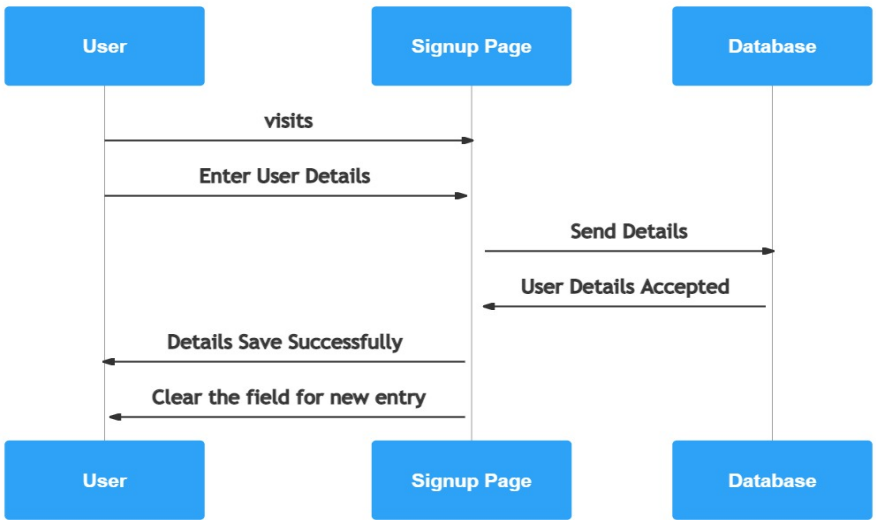


Figure 3-16 Signup Sequence Diagram

3.2.4.5.2 Login

If a User is login into the system, they must enter their login details on the login screen and press login button. The system would validate the details entered, then sends the data to the server side. User authenticated notification is generated and shown to the User, if wrong details are entered, the system would then generate an error message as shown in Figure 3-17.

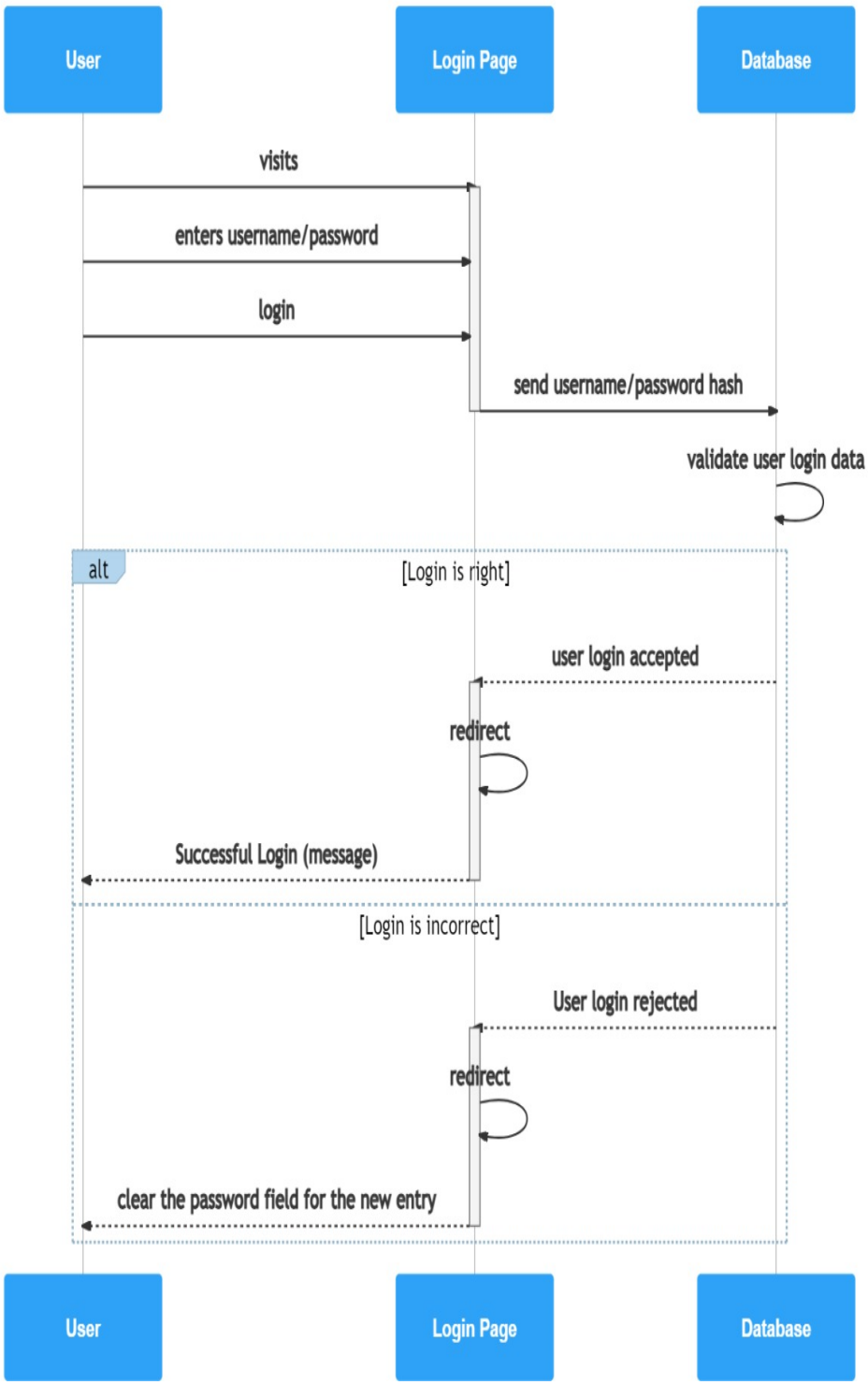


Figure 3-17. Login Screen Sequence Diagram

3.2.4.5.3 Forgot Password

If a User has forgotten their password, they must enter their email on the forgot password activity and press the send instructions option. The system would then validate the email and send a password reset link. Following this link, the User would be able to reset their password. If the details are incorrect then the system would generate an error message as shown Figure 3-18.

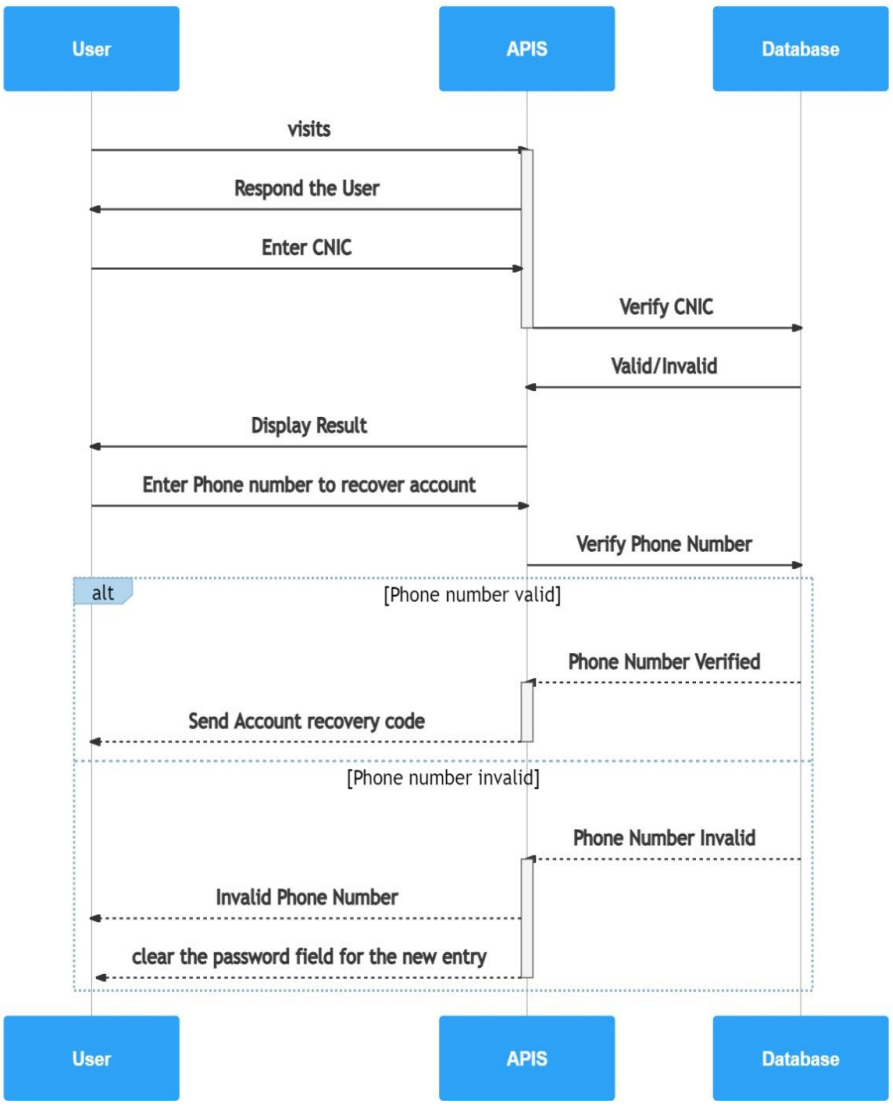


Figure 3-18. Forgot Password Sequence Diagram

3.2.4.5.4 Reset Password

If a User wants to reset their current password, then the procedure is like forgot password procedure. When the User is logged in navigate to the settings menu in the app and enter the email and password reset link will be sent otherwise the system will generate an error message as show in Figure 3-19.

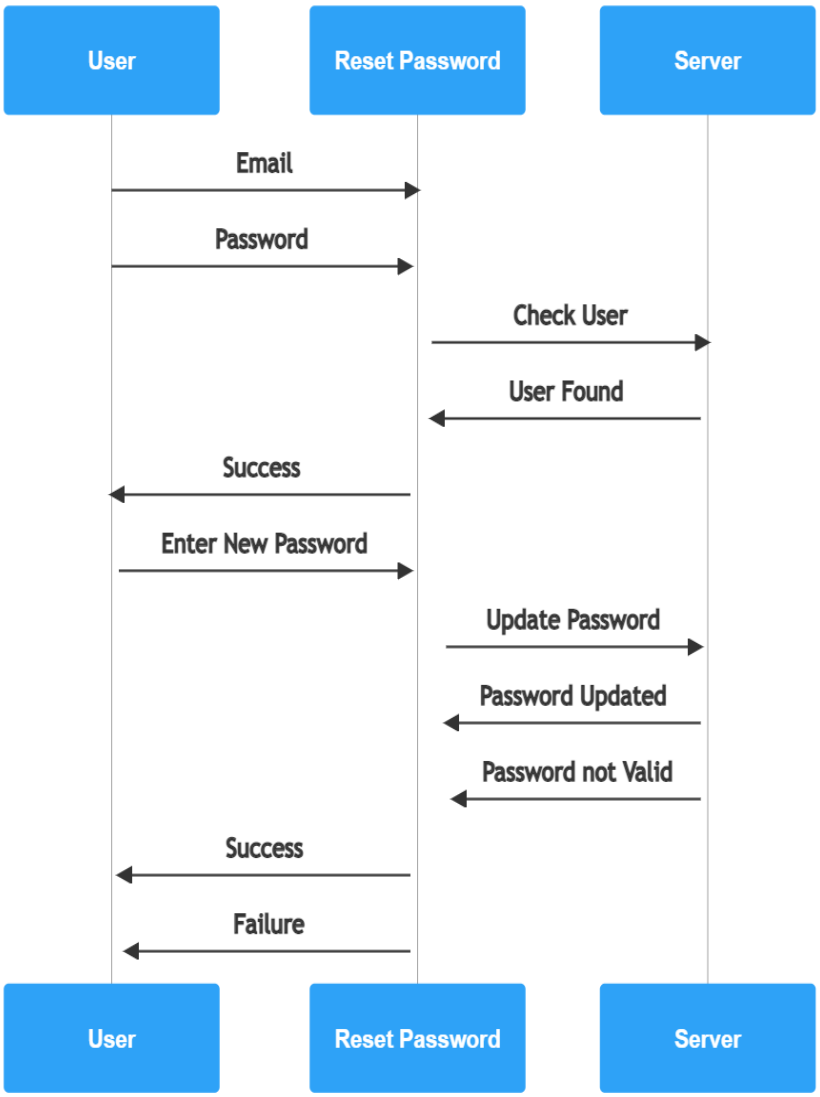


Figure 3-19. Reset Password Sequence Diagram

3.2.4.5.5 Take Picture

If a User wants to take the picture. When User is logged in, press the camera option. It will take the User to the camera. Where User can click images9 to use it for identification purposes. After Image clicked and finalized it will take the User back to image selected state as show in Figure 3-20.

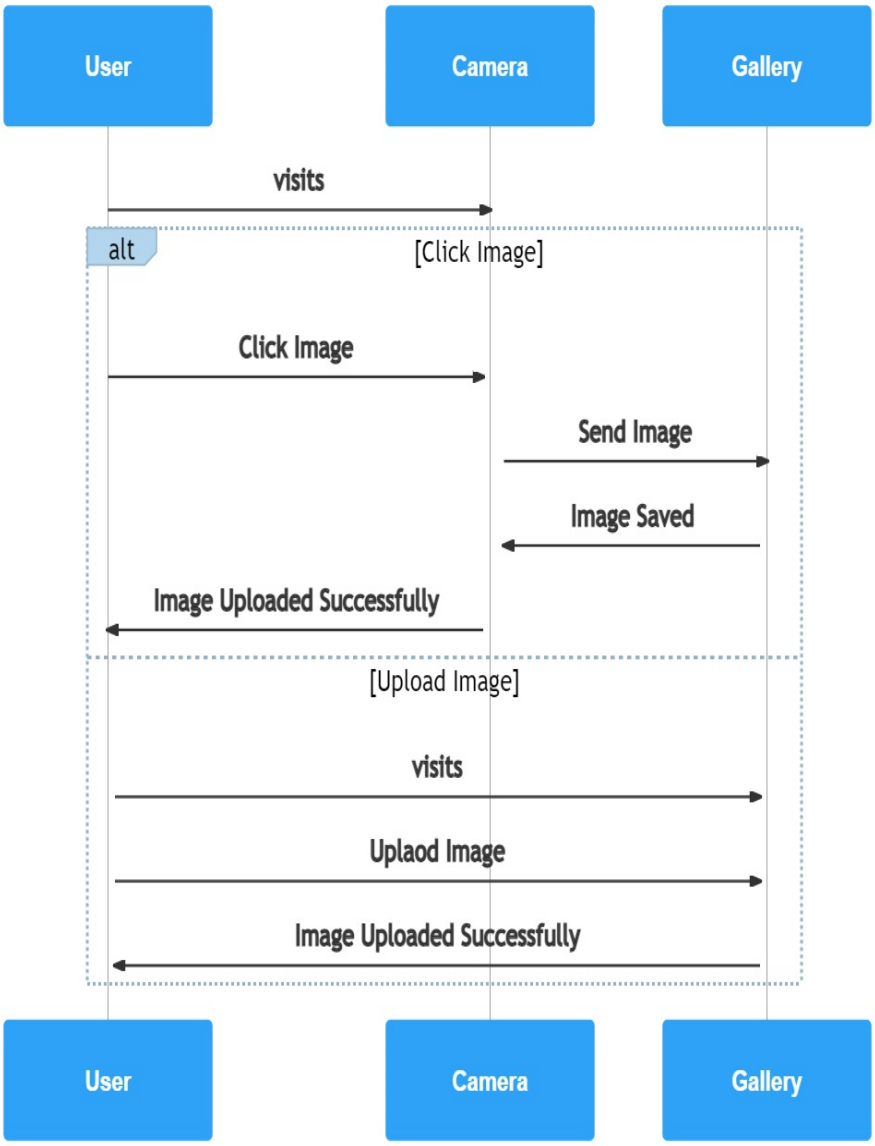


Figure 3-20 Take Picture Sequence Diagram

3.2.4.5.6 Upload Image

If User want to upload image from gallery to the application. When User is logged in, press the upload button. User is sent to the gallery activity where he can select image. After Image selected and finalized it will take the User back to image selected activity as show Figure 3-21.

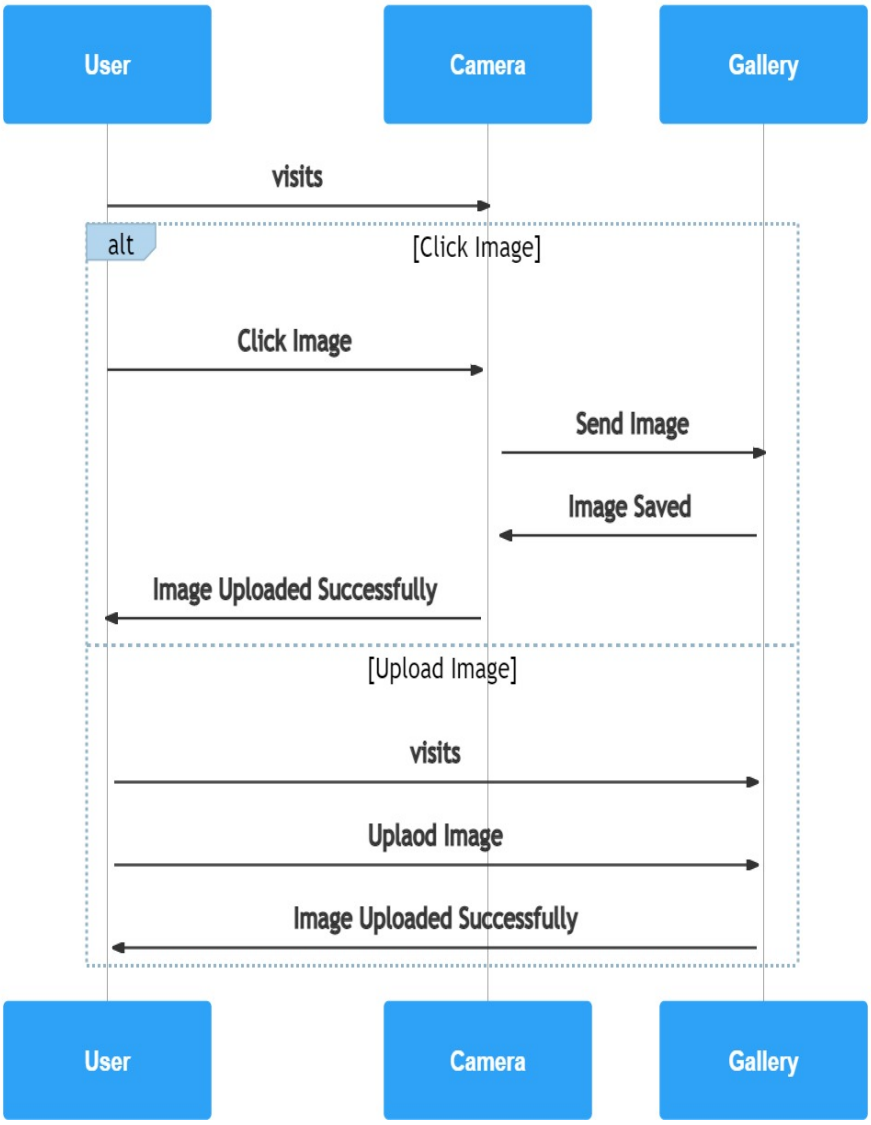


Figure 3-21 Upload Image Sequence Diagram

3.2.4.5.7 Edit Info

If the User wants to edit their information, they will enter the new information in the edit info screen in the settings menu. The system would validate the information, send it to the server side and return a positive response or if the information is invalid then the system will generate an error message as shown in Figure 3-22.

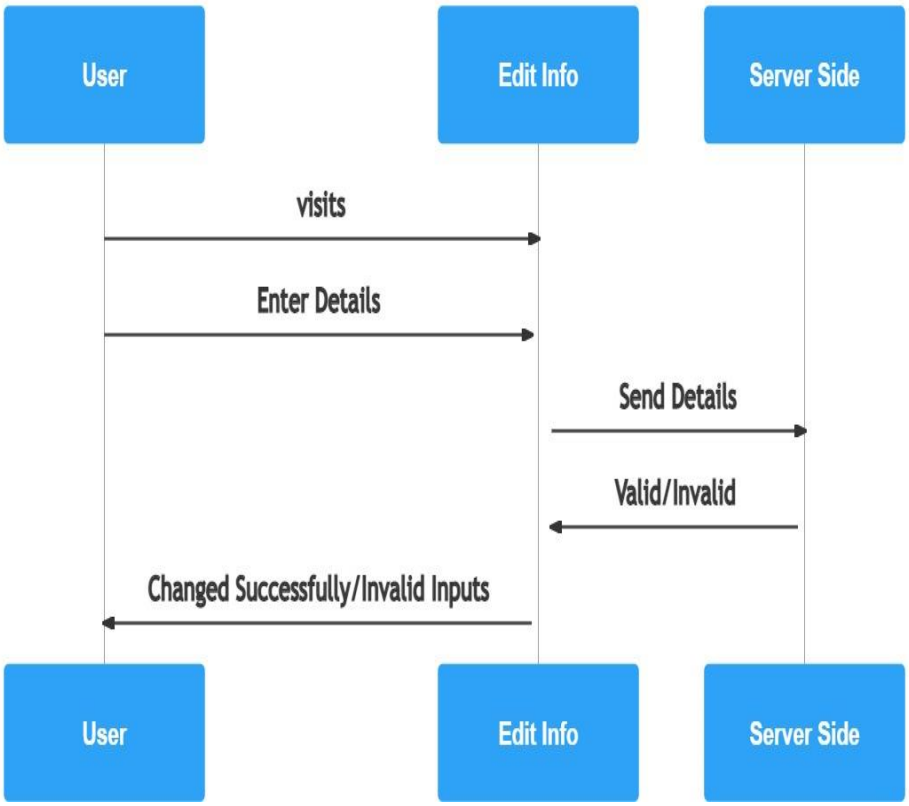


Figure 3-22. Edit Info Sequence Diagram

3.2.4.5.8 Change Image

If a User wants to change the selected image. When User is logged in and images are selected, User press the change image option. All the selected images are discarded from selected phase. User can now use upload or take picture again for use as show in Figure 3-23.

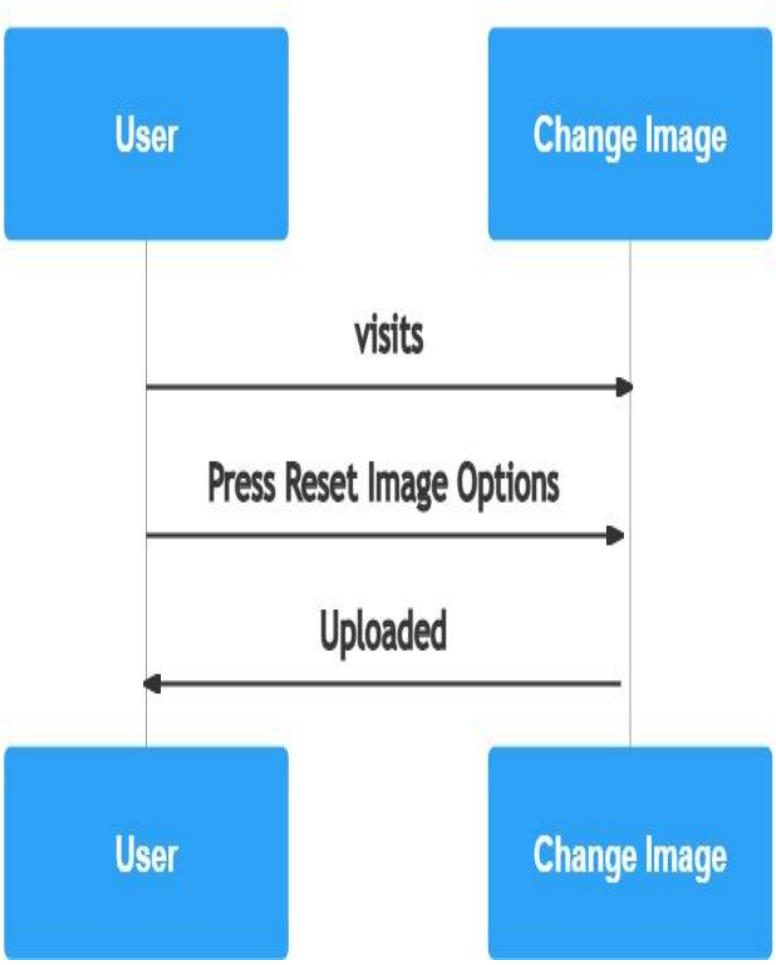


Figure 3-23 Change Image Sequence Diagram

3.2.4.5.9 View identification Result

If a User wants to view the identification result. When user is logged in and images are selected, press the view result button. System would take the selected images to CNN and all the processing is down upon it and results are generated and thrown back to the View Result tab. If invalid images are selected and used errors are thrown as show in Figure 3-24.

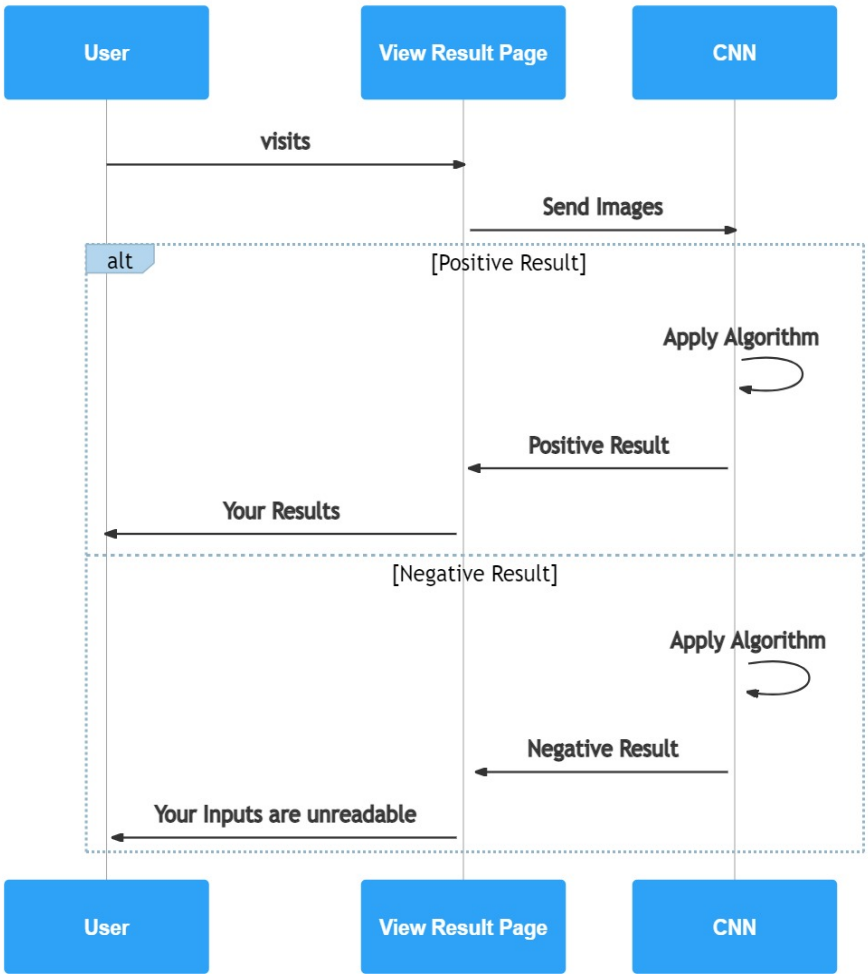


Figure 3-24 View Result Sequence Diagram

3.2.4.5.10 Save Result

If the User wants to save the result. When User is logged in and results are generated, press on the save button. System would take the User to the save image activity. Choose the desired format and press okay as shown in Figure 3-25.

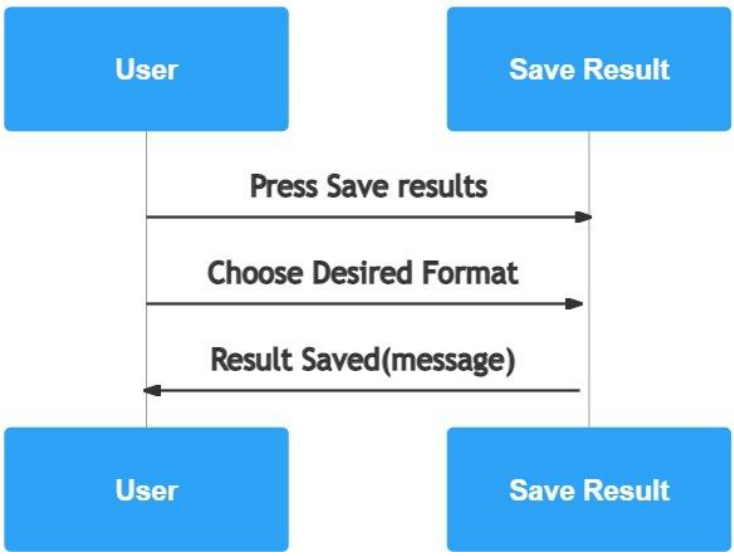


Figure 3-25. Save Result Sequence Diagram

3.3 Summary

In this chapter we discussed the proposed methodology and implementation of the model. This chapter also explained the dataset, procedure of generating the dataset. The techniques used for generating the dataset. It also discussed, pre-processing operation, train and testing split ratio and model through which identification and authentication is done. The rest of the chapter explains validation of the model and sequence diagram of the application.

CHAPTER 4

Results and Evaluation

The chapter describes result and evaluation generated by our proposed solution. The platform used for experiments is Jupyter notebook and google Colab. Google Colab is a product from Google Research. Google Colab allows anybody to write and execute python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Google Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs. This chapter also illustrate the finding of the data and experiments, data analysis and result of the model are generated and plotted using programming language python. The results are discussed in the form of accuracy, losses, and cross validation.

4.1 Evaluation measurement

The three main metrics are used to evaluate the model are accuracy, loss.

Accuracy: defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

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

Loss:

: defined as the percentage of wrong predictions for the test data. It can be calculated easily by dividing the number of wrong predictions by the number of total predictions.

$$\text{Loss} = \frac{\text{Number of wrong predictions}}{\text{Total number of predictions}}$$

4.2 Train and Test Split

The dataset must be divided into train-test split for model training and evaluation. The dataset is divided into 3 splits.

1. Train
2. Validation
3. Test

The dataset is split with following the convention used in [18]. The split rate for train is 0.5, validation 0.25 and for test is 0.25. 1000 images per class is divided into 500 images per training, 250 per validation and 250 per testing.

4.3 Experimental Setup

All the experiment is done on Jupiter notebook and google Collaboratory Pro. Collaboratory Pro is a paid service by COLAB. All the result are generated on these platforms. The device used is laptop. The specification of the laptops is:

- Processor: I5
- Generation :6
- Ram: 16GB
- Window :10

- Hard disk: 256gb

The programming language used for implementation for the proposed model is python. The version is 3.7. The implementation of the model is done only on python. The library used is Keras. The pre-processing of the dataset is done manually. From the dataset of 6000, 3000 is used for training, 1500 for validation and 1500 for testing. In terms of ratio, 0.5 is used for training, 0.25 for validation and 0.25 for test. For the evaluation metrics used are accuracy, loss, and confusion matrix

4.3.1 Model

The machine learning technique used for this project is convolution neural network. Convolution neural network is used because it is very effective and optimized when working with Visual data. Model is created in programming language python. The model is created using high level application programming interface. The version of the python language is 3.7. The library used is Keras. The epoch used are 100. The learning rate of the model is 0.0001,0.001 and 0.00025. The batch size used for model is 8,12,16. The activation function used is rectified logic unit and Softmax. The loss function used is categorical cross entropy. The optimizer used is Adam.

.

4.4 List of Experiments

Following is the list of experiments done to obtain the best accuracy of the model. .

4.4.1 Experiment 1

The purpose of this experiment was to test the dataset. Since dataset was unique, its authenticity and reliability were a question mark. To check the dataset creditability, dataset was tested on the model to solve a binary classification problem between fake emerald and real emerald. The experiment detail is provided in below table 4.1.

Table 4-1 Model and Dataset Details

Dataset	Real Emerald, fake Emerald
Number of samples	1000
Platform	Personal Computer
Problem	Binary Classification
Epochs	5
Run Time	23 Minutes

The time taken by the local computer was 23 min. Epoch time decreases with increase in the number of epochs. Total 1000 samples were used, 500 per each class. 250 images were used for training and 125 images were used for validation and rest 125 images were used for testing per class. The hyper parameter is given as below in Table 4-2.

Table 4-2 Effect of Variation of Epochs on Accuracy when Batch Size = 8 and learning rate 0.0001

Epochs	Batch Size	Learning Rate	Training Accuracy	Validation Accuracy
1	8	0.0001	82	74
2	8	0.0001	94	9
3	8	0.0001	96	094
4	8	0.0001	97	84
5	8	0.0001	97	92

The model showed huge improved after first epoch in respect of accuracy. The model showed maximum accuracy at epoch third. The creditability of the dataset was checked, and no problem was seemed to be found in the dataset. The model accuracy

showed that dataset is pretty much capable of working. More detail regarding model epoch accuracy and loss is given below in figure 4-1.

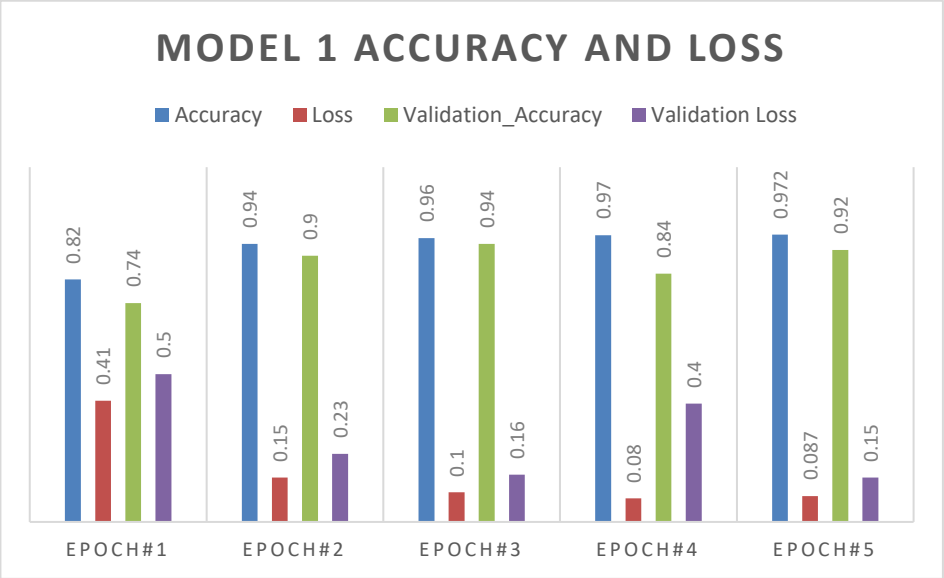


Figure 4-1. Histogram result of Experiment 1

4.4.2 Experiment 2

The purpose of this experiment was to check the model hyper parameter, accuracy, and loss estimation. Since dataset credibility was already checked in experiment 1. In experiment 2 classes were increased from 2 to 4 so accuracy. The model was same as experiment 1. Total 4 classes were used out of 6. Number of samples used for the model training were 4000. Because of the high number of images, graphics processing unit was needed. Google collaborator pro was used. Total number of epochs were 150. The time taken to completely perform the experiment is 5 hours. More detail about experiment 2 is given below in the **Table 4-3**.

Table 4-3 Model Experiment 2 Setup Details

Dataset	Emerald, Fake Emerald, Turquoise and Ruby
Number of samples	4000
Platform	Google colab pro

Epoch	150
Running Time	5 Hours

The number of samples used in this experiment were 4000. Training sample were 500, validation 250 images and testing samples were 250 images per class. The data augmentation technique was not used in our project. The total running time of the experiment was 5 hours. Average epoch completion time was 90 seconds. The model hyper parameters are defined in the table below in **Table 4-4** .

Table 4-4 Effect of Variation of Epochs on Accuracy when Batch Size = 8 and learning rate 0.0001

Parameters			Model Evaluation	
Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	8	0.0001	96.00	83.00
20	8	0.0001	99.08	95.67
30	8	0.0001	94.23	73.9
40	8	0.0001	98.92	93.33
50	8	0.0001	97.42	95.6
60	8	0.0001	98.82	96.97
70	8	0.0001	99.13	98.62
80	8	0.0001	99.36	93.33
90	8	0.0001	99.36	98.24
100	8	0.0001	99.28	98.24

The outcome of this experiment was acceptable as model was classifying images with high accuracy. During 100 epoch model accuracy was between 0.90 to 0.99. Validation and training accuracy. The testing accuracy of the model is 88%. The loss was also low during 100 epochs. Complete detail of accuracy and model during 1500 epoch is given as below 4-2.

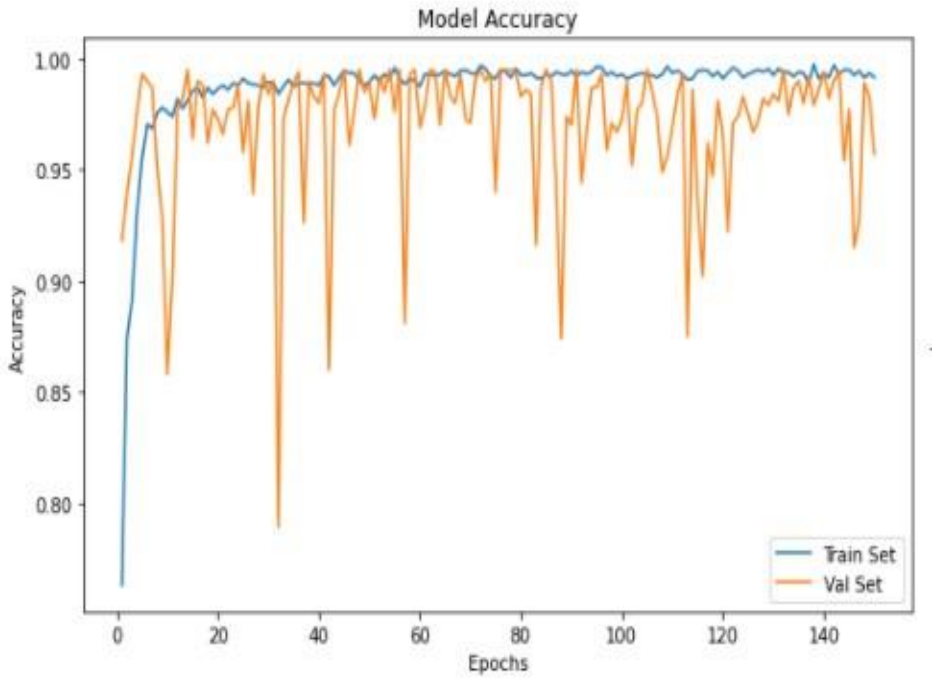


Figure 4-2. Experiment 1 accuracy Graph

Model seems overfit during middle epoch but during last epochs, model was working fine and not showing any sign of overfit or underfit. Loss was high during early epochs and loss goes down with increased in the number of the epochs. The model loss details are shown as below in figure 4.3.

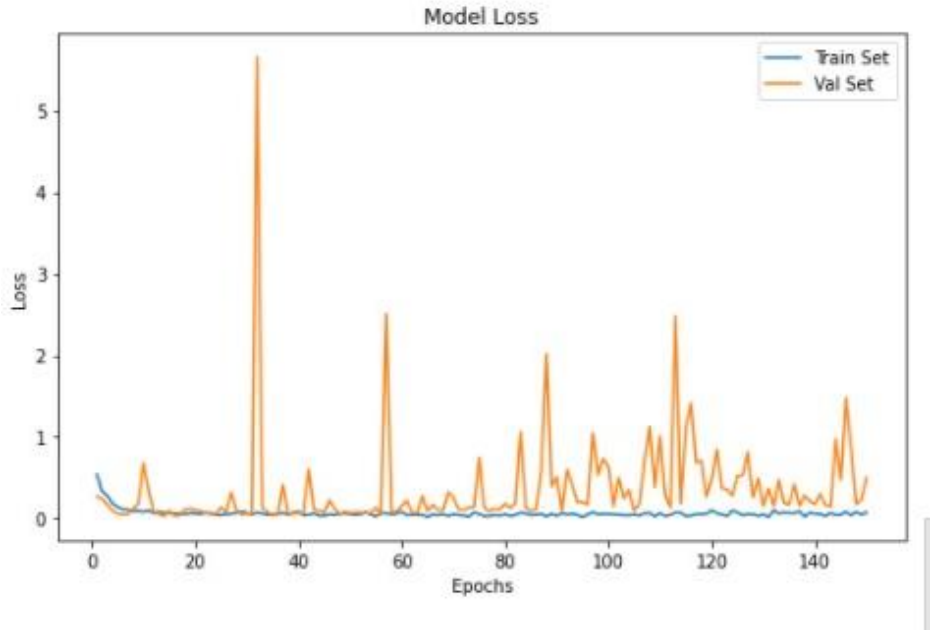


Figure 4-3. Model loss during experiment 2

4.4.3 Model Experiment 3

Highlighted row in **Table 4-5.Effect of Variation of Epochs on Accuracy when Batch Size = 16 and learning rate 0.0001** shows maximum accuracy achieved during this experiment. It is seen that model is validated as 84.67 % accurate with batch size 16 but learning rate 0.0001 having 100 number of epochs. But at these values highest accuracy is achieved at epoch 20 which is for training 99.13 and 93.0 as validation accuracy. The complete detail is in table 4-5.

Table 4-5.Effect of Variation of Epochs on Accuracy when Batch Size = 16 and learning rate 0.0001

Parameters			Model Evaluation	
Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	16	0.0001	96.24	87.24
20	16	0.0001	99.13	93.0
30	16	0.0001	98.02	91.24
40	16	0.0001	99.42	93.42

50	16	0.0001	99.73	89.65
60	16	0.0001	98.40	90.6
70	16	0.0001	100	89.50
80	16	0.0001	95.73	82.33
90	16	0.0001	89.8	75
100	16	0.0001	99.73	88.71

4.4.4 Model Experiment 4

Highlighted row in **Table 4-6** shows maximum accuracy achieved during this experiment. It is seen that model is validated as 63.7 % accurate with batch size 16 but learning rate 0.001 having 100 number of epochs. But at these values highest accuracy is achieved at epoch 10 which is for training 72.89 and 81.9 as validation accuracy. The complete detail is in table 4-6.

Table 4-6. Effect of Variation of Epochs on Accuracy when Batch Size = 16 and learning rate 0.001

Parameters			Model Evaluation	
Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	16	0.001	72.89	81.9
20	16	0.001	76.88	74.6
30	16	0.001	82.0	69.62
40	16	0.001	88.2	68.8
50	16	0.001	93.5	82.9
60	16	0.001	89.8	72.2
70	16	0.001	93.0	72.0
80	16	0.001	94.9	67.8
90	16	0.001	97.5	70.9
100	16	0.001	95.1	72.2

4.4.5 Experiment 5

Highlighted row in Table 4-7 shows maximum accuracy achieved during this experiment. It is seen that model is validated as 73.43 % accurate with batch size 16 but learning rate 0.00025 having 100 number of epochs. But at these values highest accuracy is achieved at epoch 60 which is for training 99.4 and 90 as validation accuracy. The complete detail is in table 4-7.

Table 4-7.Effect of Variation of Epochs on Accuracy when Batch Size = 16 and learning rate 0.00025

Parameters			Model Evaluation	
Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	16	0.00025	90	46
20	16	0.0025	98.3	82.7
30	16	0.00025	97	89.8
40	16	0.00025	97.6	74.8
50	16	0.00025	93.0	64.31
60	16	0.00025	99.4	90.0
70	16	0.00025	95.7	88.0
80	16	0.00025	99.4	78.6
90	16	0.00025	98.6	85.0
100	16	0.00025	99.20	81.2

4.4.6 Experiment 6

Highlighted row in Table 4-8 shows maximum accuracy achieved during this experiment. It is seen that model is validated as 95.0 % accurate with batch size 8 but learning rate 0.0001 having 100 number of epochs. But at these values highest accuracy is achieved at epoch 70 which is for training 97.13 and 97.62 as validation accuracy. The complete detail is in table 4-8.

Table 4-8. Effect of Variation of Epochs on Accuracy when Batch Size = 8 and learning rate 0.0001

Parameters			Model Evaluation	
Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	8	0.0001	99.01	98.03
20	8	0.0001	97.08	83.67
30	8	0.0001	98.23	93.9
40	8	0.0001	98.92	95.48
50	8	0.0001	99.05	96.6
60	8	0.0001	98.82	95.07
70	8	0.0001	99.13	98.65
80	8	0.0001	96.36	98.33
90	8	0.0001	98.71	95.24
100	8	0.0001	98.28	97.24

4.4.7 Experiment 7

Highlighted row in Table 4-9 shows maximum accuracy achieved during this experiment. It is seen that model is validated as 81.76 % accurate with batch size 8 but learning rate 0.0001 having 100 number of epochs. But at these values highest accuracy is achieved at epoch 10 which is for training 93.01 and 91.03 as validation accuracy. The complete detail is in table 4-9.

Table 4-9 Effect of Variation of Epochs on Accuracy when Batch Size = 8 and learning rate 0.001

Parameters			Model Evaluation	
Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	8	0.001	93.01	91.03
20	8	0.001	95.08	82.67
30	8	0.001	97.23	85.9
40	8	0.001	97.92	87.48
50	8	0.001	97.05	81.6
60	8	0.001	98.9	76.1
70	8	0.001	96.1	79.6
80	8	0.001	97.6	85.9
90	8	0.001	99.3	87.15
100	8	0.001	99.13	84.3

.

4.4.8 Experiment 8

Highlighted row in Table 4-10 shows maximum accuracy achieved during this experiment. It is seen that model is validated as 93.40 % accurate with batch size 8 but learning rate 0.00025 having 100 number of epochs. But at these values highest accuracy is achieved at epoch 40 which is for training 99.20 and 98.13 as validation accuracy. The complete detail is in table 4-10.

Table 4-10. Effect of Variation of Epochs on Accuracy when Batch Size = 8 and learning rate 0.00025

Parameters			Model Evaluation	
Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	8	0.00025	98.01	96.03
20	8	0.00025	99.08	97.67
30	8	0.00025	99.23	97.9
40	8	0.00025	99.20	98.13
50	8	0.00025	99.13	97.53
60	8	0.00025	99.0	97.12
70	8	0.00025	98.65	94.9
80	8	0.00025	97.7	92.01
90	8	0.00025	99.3	97.66
100	8	0.00025	99.5	97.86

4.5 User manual

Gemo is an android application created to use the trained model for gemstone authentication. The android jellybeans are minimum required to run this application. It is advised to use camera of more than 50 mega pixel. The background reaccommodated for the images is white. User should capture gemstone image in the natural shape, it is advised to avoid flash effect on the surface of gemstone for better and accurate results. The accurate results will only be provided if image of gemstones included in the scope of the project are used. There is total 8 screen of the gemo application. These screens are:

- Login
- Sign up
- Forget Password
- Reset password

- Identification
- Take image
- Select image
- View result

Each of these screens are explained in detail in the next subsections.

4.5.1 Login

If a User wants to login into the gemo application, they must enter their login details on the login screen—email, password. Then enter press login button. The system would validate the details entered, then sends the data to the server side. User authenticated notification is generated and shown to the User, if wrong details are entered, the system would then generate an error message. The database used is firebase. The UI of the login screen is shown in Figure 4-4.

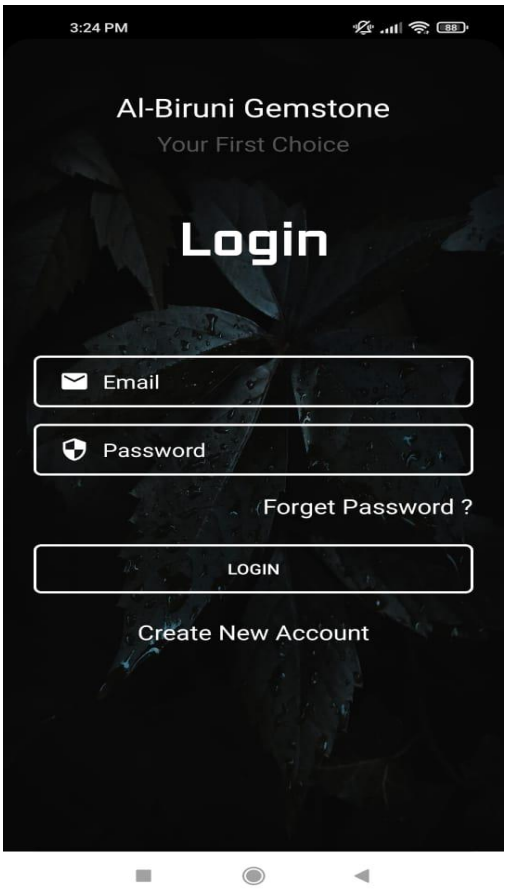


Figure 4-4. Login Screen of Gemo application

4.5.2 Register

If a new User wants registered into the application, they must enter the register details on the signup activity and press register button. The system would validate the details entered, then sends the data to the server side. Further user created notification is generated and shown to the User, if wrong details are entered, the system would then generate an error message. The UI of the register screen is shown in Figure 4-5.

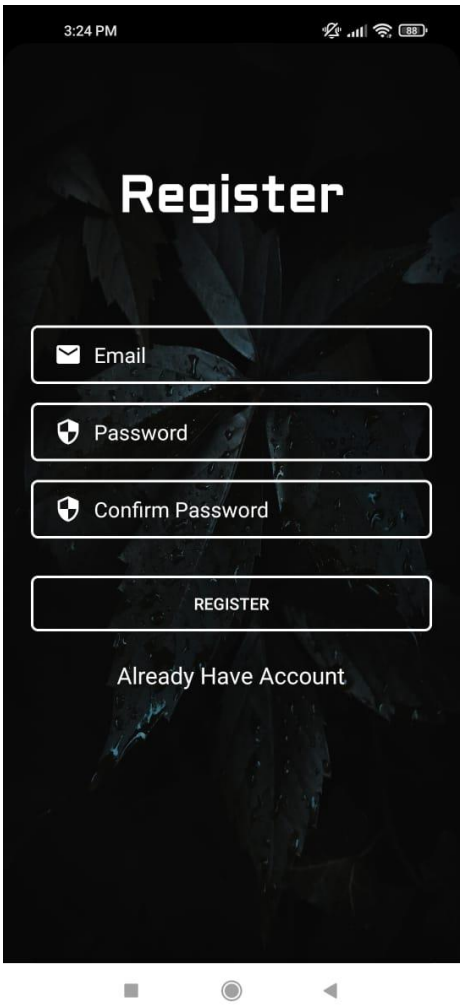


Figure 4-5 Register Screen of Gemo application

4.5.3 Forget Password

If the User forget his password, he will press forget password. Then forget password screen appear. The user will enter his email. The system would validate the information, send it to the server side and return a positive response or if the information is invalid then the system will generate an error message. The UI of the forget password screen in Figure 4-6.

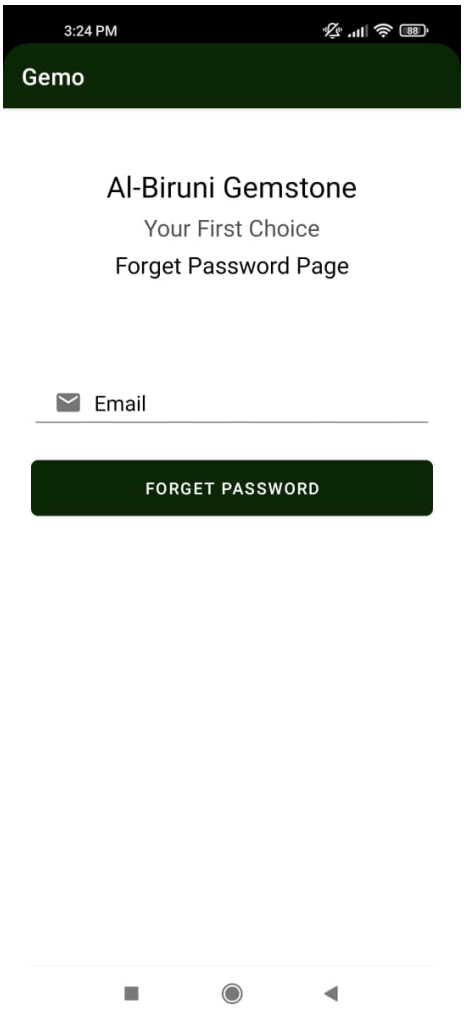


Figure 4-6 forget password screen of gemo

4.5.4 Reset Password

This screen is for the purpose to reset the password. For password reset, user will select reset password option for dashboard. Then reset password screen will appear, use enter old password and then new password. After entering both new and old password, press change password. If the provided information is correct, new password will be set for user. The UI of the reset password shown in Figure 4-7.

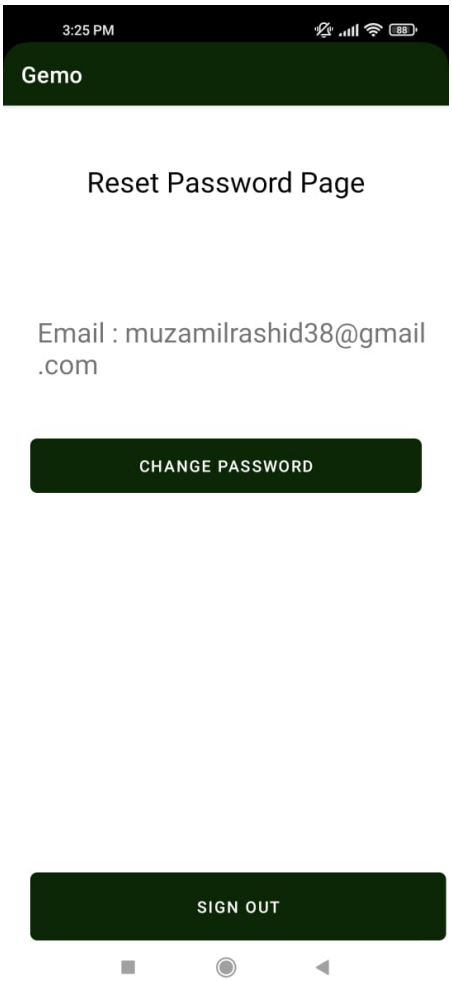


Figure 4-7 Reset Password screen of gemo

4.5.5 Take Image

Take image screen will help the user to take the live images of gemstone. User will first login and after successful login. If user select take image option inside identification feature, application will detect camera or check, or permission allowed for the camera. If both conditions fulfilled, user would get the access to camera and camera screen will appear. On the contradiction side error will be displayed. Upon camera access user can click images and use it in application. The UI of the Take image screen shown in Figure 4-8.

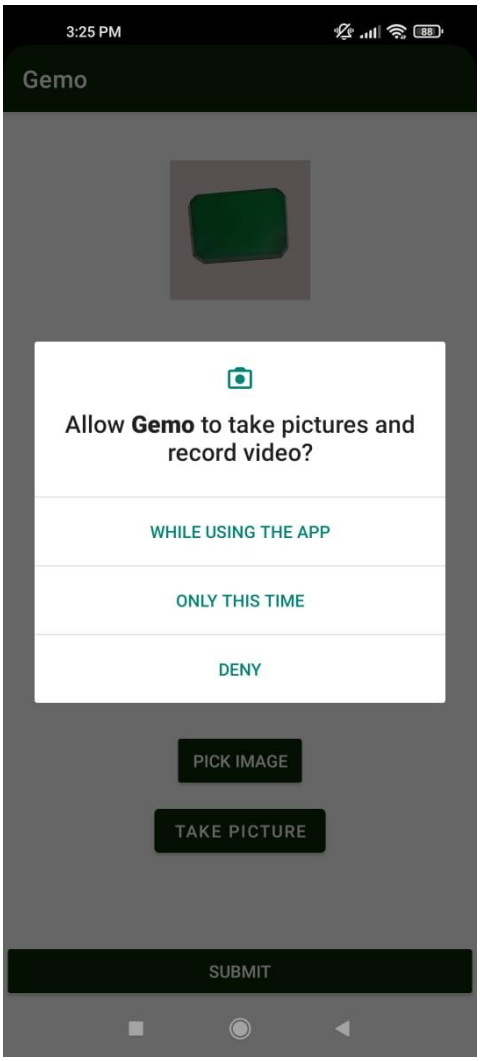


Figure 4-8 Take Image screen of gemo application

4.5.6 Select Image

Select image screen will help the user to upload images of gemstone. User will first login and after successful login. Application will ask for permission allowed for the gallery, if not granted. If both conditions fulfilled, user would get the access to gallery and gallery screen will appear. On the contradiction side error will be displayed. Upon gallery access user can upload images and use it in application. The UI of the select image is shown in Figure 4-9.

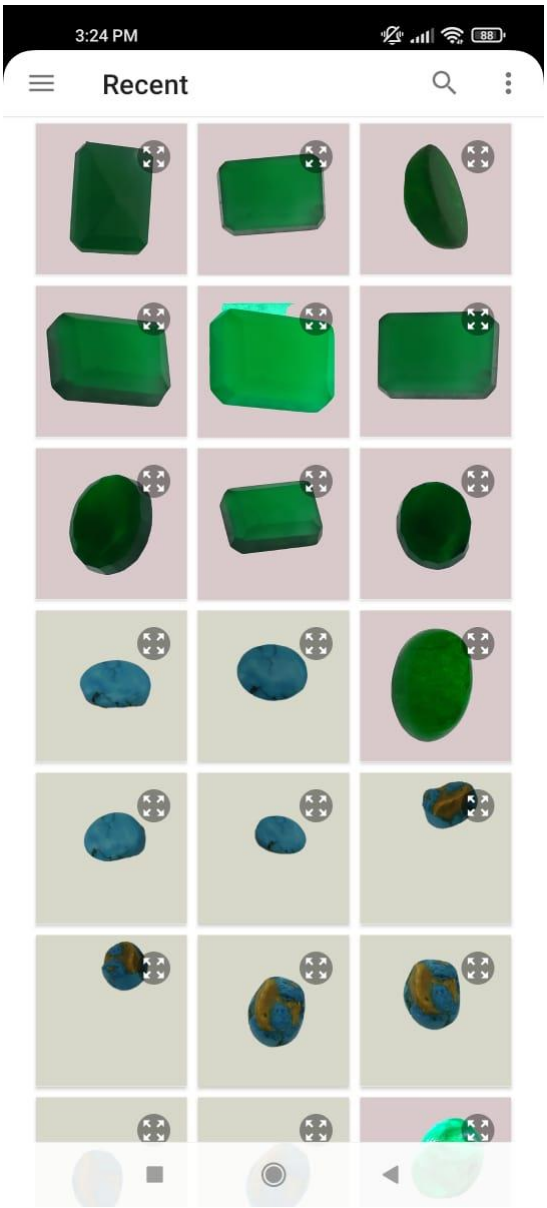


Figure 4-9 Select Image screen of gemo application

4.5.7 Identification of gemstone

This screen will provide the user identification of gemstone feature. The user can select this feature from dashboard screen. The identification feature includes take image and select image as its sub feature. The user will select image or take image feature to fit the screen image view. On submit button press, the image is sent to model in back-end of this screen, identification process is done. The results are

generated and send to view result screen. The UI of identification of gemstone screen is shown in Figure 4-10.

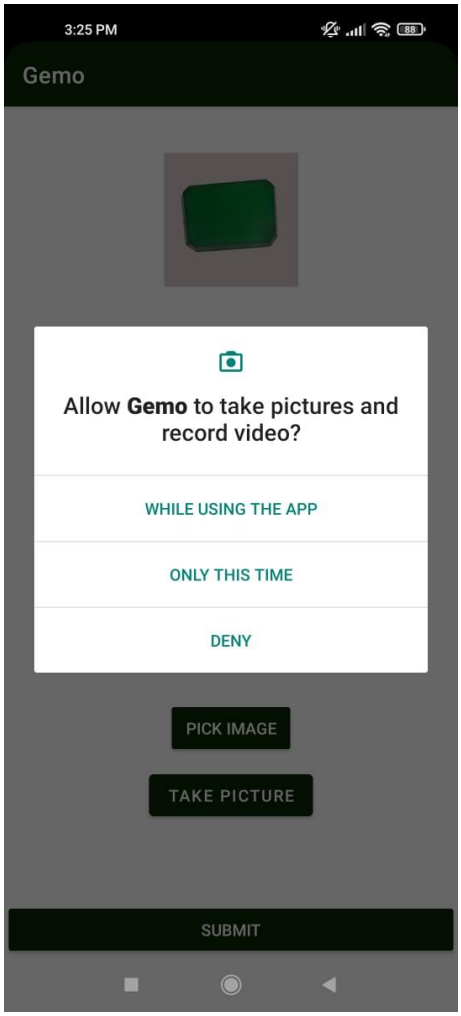


Figure 4-10 Gemo application identification gemstone screen

4.5.8 Result

The purpose of the screen is to display the prediction result of images. The result is predicted by the model used in identification screen. The prediction result is identified by the identifier used for prediction name. The prediction is sorted by the latest to old. This screen also allowed the user to remove prediction by press long tap on prediction. The UI of the result screen in shown in Figure 4-11.

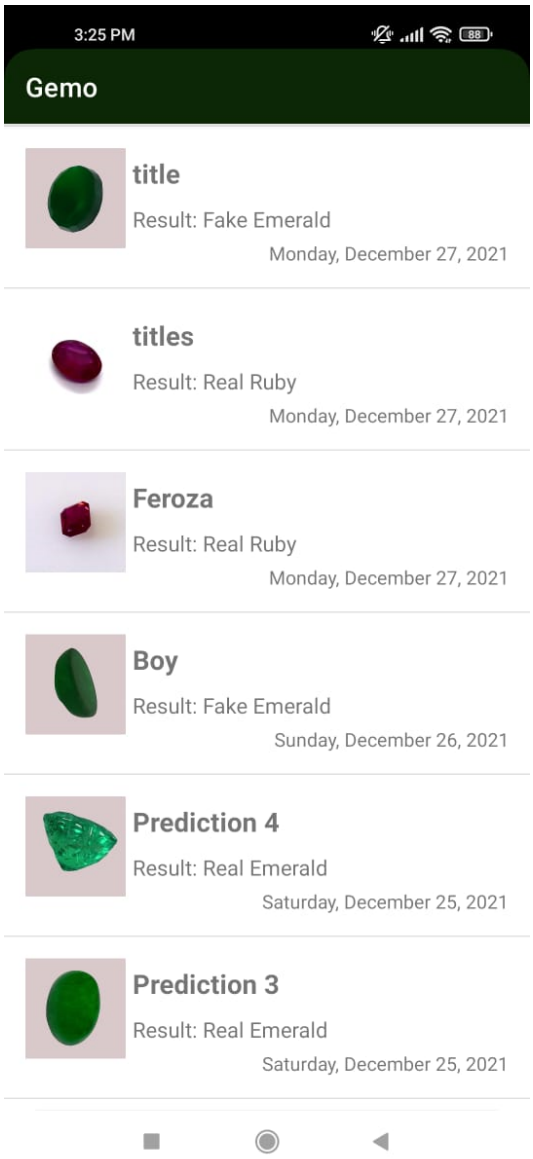


Figure 4-11 Result screen of the gemo application

4.6 Summary

This chapter explained the model experiment environment setup. The platform used were google Colab and Jupyter notebook. Three learning rates 0.0001,0.001 and 0.00025 were used. The batch size was 8. Total number of experiments were 6. The highest accuracy was 95%.

CHAPTER 5

RESULTS AND DISCUSSION

In this chapter results obtained from experiments above which are elaborated in chapter 4 are discussed. Proposed model is fine tunned using some parameters such as epochs and learning rate using machine learning techniques. For the experiment generated dataset which is already discussed in detail in Phase 2. Model is experimented in three phases first training, validation and then testing based on accuracy.

5.1 Results Summary

To perform the experiment python is used as a programming language along with Google Colab to use free GPU and cloud services with easy-to-use integrated machine learning libraries in a jupyter notebook environment. For this purpose, dataset is split into three parts—training, validation, testing. The training with ratio of 50% for training and 25% for Validation and 25% for Testing. The Experiment were performed by fine tuning the model through its parameters. At last k-fold cross validation is implemented to validate the model.

From experimentation of model on custom dataset in it is clearly shown that best training accuracy which is 99.50% is achieved when batch size is 8, Validation accuracy for this experiment is 98.67% at epoch 80. Best Validation accuracy of the model which is 98.67 is achieved when batch size 8, Testing accuracy for this experiment is 95.0 % at 100 epochs.

Results can better be observed by comparing accuracy with hyperparameters used for tuning the model. When accuracy is compared with epochs, batch size and learning rate. It is easy to evaluate and show on which stages model has performed at its best.

5.1.1 . Epoch vs Accuracy

Number of epoch when compared with accuracy gives us insight at the performance of the model. We can see in Table 5-1 as the number of epochs are increased it has little effect on Accuracy achieved. Best accuracy was achieved on 70 epochs which is 98.65 % where learning rate is 0.0001 and Batch Size is 8. Hyperparameters of Activation function is SoftMax, and Optimizer is Adam who is kept constant.

Table 5-1 Effect of Epochs on Accuracy when Optimizer = Adam and SoftMax

Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	8	0.0001	99.01	98.03
20	8	0.00025	99.08	97.67
30	8	0.00025	99.23	97.9
40	8	0.00025	99.20	98.13
50	8	0.00025	99.13	97.53
60	8	0.00025	99.0	97.12
70	**8**	**0.0001**	**99.17**	**98.65**
80	8	0.0001	96.36	98.33
90	8	0.00025	99.3	97.66
100	8	0.00025	99.5	97.86

5.1.2 Batch Size vs Accuracy

Batch size when compared with accuracy gives us insight at the performance of the model. We can see from Table 5-2 that there is a slight increase in accuracy when the

batch sizes have increased. Best accuracy was achieved on 70 epochs which is 98.65% where learning rate is 0.0001 and Batch Size is 8. Hyperparameters of Activation function is SoftMax, and Optimizer is Adam who is kept constant

Table 5-2.Effect of Batch Size on Accuracy when Optimizer = Adam and SoftMax

Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
20	12	0.00025	99.08	97.67
20	16	0.0001	99.13	93.0
70	**8**	**0.0001**	**99.17**	**98.65**

5.1.3 Learning Rate vs Accuracy

Learning rate when compared with accuracy gives us insight at the performance of the model. We can see as the learning rate decreases Accuracy has increased. From previous tables it can be observed that when accuracy was set to 0.0001 the model has given better accuracy than other two. This learning rate was most influential in obtaining best accuracies of model. Best accuracy was achieved on 80 epochs which is 98.33% where learning rate is 0.0001 and Batch Size is 20. Hyperparameters of Activation function is SoftMax, and Optimizer is Adam who is kept constant.

Table 5-3.: Effect of Learning Rate on Accuracy when Optimizer = Adam and SoftMax

Epoch	Batch size	Learning rate	Accuracy	Validation Accuracy
10	8	0.001	93.01	91.03
20	12	0.00025	99.08	97.67
70	**8**	**0.0001**	**99.17**	**98.65**

5.1.4 Overall Result

From the sections above it has been seen that best optimal results in case of both accuracy and validation accuracy is achieved from batch size 8 which are 99.17 training accuracy and 98.65 validation accuracy at 70 epochs of the experiment. It is observed from above sections that learning rate had the most impact on accuracy of model. Accuracy have also improved when batch size and epochs has increased. It can be concluded that presented model works best on our dataset when parameters such as epoch, batch size and learning rate are 70, 8 and 0.0001. After the experiments best model is integrated with the application which is developed using

- Python
- Tensor flow API
- Keras
- Android Studio

Jupiter is used as a python IDE and Android studio is used to developed Android Application. Keras is used to load the model in the application code using .h5 file which is saved while experiments. Tensorlite API is used to integrate the model with android application. The .h5 file contains the detail of the model which is fine stunned by changing values of some of the parameters such as epoch, learning rate to find optimal results by using proposed model on dataset of gemstones.

5.2 Authentication of gemstones using Application

Gemstone authentication is done using android application named gemo. First user will login if user is already register or signup in the application. Data of users Signup and login is saved in database. Database which is used for application is firebase. After login application will offer three main features—identification, authentication, and view results. The identification and authentication will be done using select image and camera clicked image. The detail of these features is explained in Software Requirement Specification/Literature Review in chapter 2.

Following are details and results related to experiment for the identification of gemstones using application.

Location: Home
Data: Images from gallery
Equipment: Smart Phone
Platform: Gemo application
Problem: Authentication of Emerald stone
Runtime: Few seconds
Best Detection time: 0.5 milliseconds

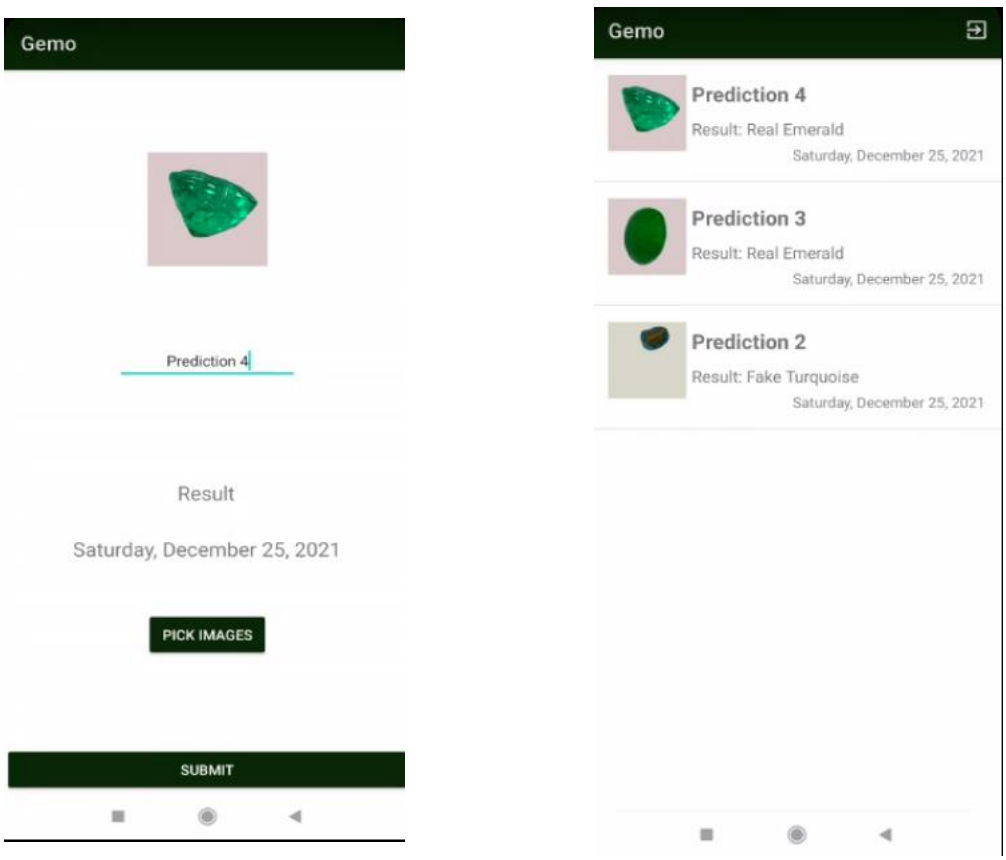


Figure 5-1 Emerld image experiments using gemo application

5.3 K-Fold Cross Validation

K-Fold Cross-validation is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modelling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

To get the best out of the model, best result and to avoid biasness in the proposed model K-Fold cross validation is applied. To get the optimal result Cross validation used is CV-5. The accuracy achieved after cross validation was approximately 94.22% and loss was 0.0589.

5.4 Summary

In this chapter model results and experiments were discussed. The metrics used for evaluation are explained. The results of the model before K Fold cross validation and improve after using K fold cross validation.

CHAPTER 6

Conclusion and Future Work

This is final chapter of the Project and will be compromised of conclusion and research work. This study can be used to create an authentication gemstone system at commercial level.

6.1 Conclusion

The gemstones are one of most traded products around the Globe.it is impossible for layman to distinguish between real or fake stones. There are many kinds of gemstones. Boundary between them is very narrow. Even specialized people of gemstone stone can get confused. Current solution used for this problem is time consuming and expensive. As lowest identification fee is 1000 through laboratory and takes 1-2 days. The proposed solution authenticates between real and fake gemstones of three kind—emerald, turquoise, ruby. For the completion of the solution, literature review was carried out and literature gap was found. The benchmark of this research to improve the accuracy achieved in [9]. The accuracy achieved in this study was 90.66%.

There was no proper dataset available for gemstones. There was only one dataset available on Kaggle. The dataset was not detailed enough. There was no data regarding fake gemstones. Dataset generation was essential for the project. A dataset of 6000 images consisting of 6 classes was generated.

. For the optimal result model three hyper parameter—epoch, batch size and learning rate, learning rate used were 0.001, 0.0001, and 0.00025. The number epochs were 100. The batch sizes were 8, 12, and 16. The optimal model was achieved at learning

rate 0.0001, batch size of 8, epoch number 70. The accuracy achieved is 95.0%. Adam. The final accuracy of the model is 95.0%. For the validation purpose K-Fold cross validation was used and accuracy was 94.22%. There is no significant difference between Kfold and model accuracy , model is neither under fit or over fit.

For ease of access, model is integrated into android application. The android application provides two main feature identification and authentication. The application take input from the user in the form of images. Images can be taken from camera or selected from gallery. The result will be saved inside the application. The aim of the study was successfully achieved. The accuracy achieved by the model is 95.0%.

6.2 Research Objective Discussion

This section would provide abstract detail regarding research objective and its related discussion. Below are the research objectives.

1. Objective 1: Identification of gemstone

   First objective was to create an identification system using deep learning technique. This objective was achieved because of the dataset generation and generated data was large enough to makes it possible.

2. Objective 2: Authentication of Gemstone

   Second objective and arguable the most important objective of the research was to create an authentication system using machine technique. There were 6 classes, pair of them interrelated with each other. Fake turquoise, emerald, ruby were the fake classes and genuine are real classes of these stones. The proposed model successfully distinguished between these classes. The accuracy achieved is 95%.

3. Objective 3 Filling the research gap

In the section 2.9.3 Research Gap, it is explained what this study is trying to achieve. The objective was to create a model for 3 stones and achieving accuracy more than 90 %. This objective was also successfully achieved.

4. Objective 4 Generation of the dataset

   In the section Benchmark Dataset 3.2.2.1, it is mentioned that there is no standard or proper dataset available to carry out the study. The generation of the dataset was also included and was an objective of the study. This objective was also achieved as 6000 images of gemstones was generated.

6.3 Future Work

The future is to create a large dataset which cover more gemstones in its scope and more in samples. The transfer learning of model to pretrained to increase model reliability and accuracy. The hyper parameter tuning to achieve the best setting for the proposed model to achieve the best accuracy and improvement in the proposed model.

REFERENCES

[1] "Strategic Plan for Pakistan ' s Gems and Jewelry Industry," 2006.

[2] L. Leon-Reina, J. M. Compana, Á. G. De la Torre, R. Moreno, L. E. Ochando, and M. A. G. Aranda, "Powder diffraction analysis of gemstone inclusions," *Powder Diffraction*, vol. 26, no. 1, pp. 48–52, 2011, doi: 10.1154/1.3552672.

[3] W. Aizuddin, W. Razali, N. Atiqah, B. Baharudeen, F. Najihah, and B. Mat, "Gemstone identification using fluorescence technique," vol. 3, no. September, pp. 73–78, 2020.

[4] J. Jehlička, A. Culka, M. Baštová, P. Bašta, and J. Kuntos, "The Ring Monstrance from the Loreto treasury in Prague: Handheld Raman spectrometer for identification of gemstones," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2082, 2016, doi: 10.1098/rsta.2016.0042.

[5] S. Saeseaw, N. D. Renfro, A. C. Palke, Z. Sun, and S. F. McClure, "Geographic origin determination of emerald," *Gems and Gemology*, vol. 55, no. 4, pp. 614–646, 2019, doi: 10.5741/GEMS.55.4.614.

[6] E. Fritsch and C. M. Stockton, "Infrared spectroscopy in gem identification.," *Gems & Gemology*, vol. 23, no. Spring, pp. 18–26, 1987, doi: 10.5741/gems.23.1.18.

[7] L. Leon-Reina, J. M. Compana, Á. G. De la Torre, R. Moreno, L. E. Ochando, and M. A. G. Aranda, "Powder diffraction analysis of gemstone inclusions," *Powder Diffraction*, vol. 26, no. 1, pp. 48–52, 2011, doi: 10.1154/1.3552672.

[8] W. Aizuddin, W. Razali, N. Atiqah, B. Baharudeen, F. Najihah, and B. Mat, "Gemstone identification using fluorescence technique," vol. 3, no. September, pp. 73–78, 2020.

[9] I. Maula, V. Amrizal, A. H. Setianingrum, and N. Hakiem, "Development of a Gemstone Type Identification System Based on HSV Space Colour Using an Artificial Neural Network Back Propagation Algorithm," vol. 149, no. Icosat 2017, pp. 104–109, 2018, doi: 10.2991/icosat-17.2018.24.

[10] A. Ostreika, M. Pivoras, A. Misevičius, T. Skersys, and L. Paulauskas, "Classification of objects by shape applied to amber gemstone classification,"

Applied Sciences (Switzerland), vol. 11, no. 3, pp. 1–14, 2021, doi: 10.3390/app11031024.

[11] J. T. Qiu, L. Qiu, and H. X. Mu, "Feasibility of gem identification using reflectance spectra coupled with artificial intelligence," *Spectroscopy Letters*, vol. 52, no. 9, pp. 520–532, 2019, doi: 10.1080/00387010.2019.1671873.

[12] S. Sinkevicius, A. Lipnickas, and K. Rimkus, "Amber gemstones sorting by colour," *Elektronika ir Elektrotechnika*, vol. 23, no. 2, pp. 10–14, 2017, doi: 10.5755/j01.eie.23.2.17993.

[13] D. Wang, L. Bischof, R. Lagerstrom, V. Hilsenstein, A. Hornabrook, and G. Hornabrook, "Automated Opal Grading by Imaging and Statistical Learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 2, pp. 185–201, 2016, doi: 10.1109/TSMC.2015.2427776.

[14] S. Sinkevicius, A. Lipnickas, and K. Rimkus, "Multiclass amber gemstones classification with various segmentation and committee strategies," *Proceedings of the 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems, IDAACS 2013*, vol. 1, no. September, pp. 304–308, 2013, doi: 10.1109/IDAACS.2013.6662694.

[15] "Machine Learning: What it is and why it matters | SAS." https://www.sas.com/en_us/insights/analytics/machine-learning.html#:~:text=Machine learning is a method,decisions with minimal human intervention. (accessed April 3, 2021) (accessed Jun. 25, 2021).

[16] "Emerald. [Online]. Available: https://www.kuwaitgems.com/emerald/. [Accessed: 26-Dec-2021]." https://www.kuwaitgems.com/emerald/ (accessed Dec. 27, 2021).

[17] "'Gem auctions online - verified gemstone dealers - gemstones for sale,' Gem Rock Auctions. [Online]. Available: https://www.gemrockauctions.com/. [Accessed: 26-Dec-2021]." https://www.gemrockauctions.com/ (accessed Dec. 27, 2021).

[18] F. Chollet, *M A N N I N G*.