AN AI APPROACH ON VEHICLE THEFT DETECTION IN PARKING AREAS WITH AN ALERT MESSAGE

A PROJECT REPORT

Submitted by

GOWTHAM .E [211419104086]

HARI LINGA PRADEEP .M [211419104090]

MD. AAMIR [211419104165]

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)

APRIL 2023

PANIMALAR ENGINEERING COLLEGE

(An Autonmous Institution, Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project report **"AN AI APPROACH ON VEHICLE THEFT DETECTION IN PARKING AREAS WITH AN ALERT MESSAGE"** is the bonafide work of "**GOWTHAM E (211419104086) HARI LINGA PRADEEP M (211419104090) MD AAMIR (211419104165)"** who carried out the project work under my supervision.

SIGNATURE

Dr.L.JABASHEELA,M.E.,Ph.D.,
HEAD OF THE DEPARTMENT

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE,
NASARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

SIGNATURE

Dr.N.PUGHAZENDI,M.E.,Ph.D.,
PROFESSOR

DEPARTMENT OF CSE,
PANIMALAR ENGINEERING COLLEGE,
NASARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

Certified that the above candidates was examined in the End Semester Project Viva-Voce Examination held on.....

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION BY THE STUDENT

We "**GOWTHAM E (211419104086) HARI LINGA PRADEEP M (211419104090) MD AAMIR (211419104165)"** hereby declare that this project report titled "A**N AI APPROACH ON VEHICLE THEFT DETECTION IN PARKING AREAS WITH AN ALERT MESSAGE"** under the guidance of **Dr. N. PUGHAZENDI,M.E.,Ph.D.,** is the orginial work done by us and we have not plagiarized or submitted to any other degree in any university by us.

GOWTHAM E

HARI LINGA PRADEEP M

MD AAMIR

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our respected Secretary and Correspondent **Dr.P.CHINNADURAI, M.A., Ph.D.** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We express our sincere thanks to our beloved Directors **Tmt.C.VIJAYARAJESWARI**, **Dr.C.SAKTHI KUMAR, M.E.,Ph.D** and **Dr.SARANYASREE SAKTHI KUMAR B.E.,M.B.A.,Ph.D.,** for providing us with the necessary facilities to undertake this project.

We also express our gratitude to our Principal **Dr.K.MANI, M.E., Ph.D.** who facilitated us in completing the project.

We thank the Head of the CSE Department, **Dr. L.JABASHEELA , M.E.,Ph.D.,** for the support extended throughout the project.

We would like to thank my guide and co-ordinator **Dr. N. PUGHAZENDI, M.E.,Ph.D.,** and all the faculty members of the Department of CSE for their advice and encouragement for the successful completion of the project.

GOWTHAM E

HARI LINGA PRADEEP M

MD AAMIR

ABSTRACT

Two wheeler Theft is a common criminal activity that is prevailing over the years and is increasing day by day. To tackle this problem many surveillance systems have been introduced in the market. Some are simply based on video surveillance monitored by a human while some are AI-based capable of detecting suspicious activity and raising an alarm. However, none of them are intelligent enough to identify what kind of suspicious activity is being carried out and what kind of protective measures should be taken in real-time. The background of this research is to build a system that will be able to help reduce crime, especially not the crime of motor vehicle theft by using the concept of facial recognition on every motorized vehicle.. The issue addressed in this project is how to lower the rate of auto theft utilising the system technique; hence, by creating this system, it will be able to assist the police in lowering the crime of auto theft. This project's goal is to show how to develop a system that can lessen motor vehicle crime, particularly motor vehicle theft, so that drivers can feel confident leaving their powered vehicle wherever. Consequently, installing a facial recognition system on two wheelers will help lower crime rates and provide drivers more confidence and also it provides the additional information about the theft , if it is parked in particular location if any crime happens it send some information to the owner. And also the crime rate is also reduced due to facial recognition system we can easily find the theft person. And the datas of the owner is already register in the application and we can detect the known and the unknown person who are all accessing the vehicle.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS

GBTD	Gradient Boosting Theft Detector
GBC	Gradient Boosting Classifiers
ETD	Electricity Theft Detection
RFE	Recursive Feature Elimination
WFI	Weighted Feature Importance
SAE	Stacked Auto Encoder
GPS	Global Positioning System
XML	Extensible Markup Language

1.INTRODUCTION

1.INTRODUCTION

The high number of cases of motor vehicle theft, especially motorcycles, is a daily problem that still needs to be solved. One of the ways to secure motorcycles from theft. One of the smart security system technologies in the field of transportation is an intelligent security system for motorized vehicles. Several methods of motorized vehicle security systems that are currently being used include facial recognition or facial recognition applied to the vehicle, one of the technologies that has a high enough accuracy where the user must first register a photo of the user's face that must be stored in the database .

1.1 PROBLEM DEFINITION

One of feature is to provide a notification system when it is not the owner of the vehicle who is driving the motorcycle. The purpose of building a security system, Increase security on this motorcycle, Minimize the possibility of vehicle theft, Increase a sense of security to the rider . The method used in this research is to use the literature review method based on previous projects so that this project makes the latest research with novelty from research based on previous research, therefore by reading a lot of previous research it will find problems and future research .

The problem raised in this study is how to reduce crime, by early detection of vehicle owners. Therefore, with the system offered in this study, it will be able to recognize who is the real owner of the vehicle that will be tested on the owner's facial recognition. The purpose of this study is how to make a proposed system that can make vehicles recognize their owners. Therefore, with a face detection system for the owner of the vehicle, it will be able to reduce crime because it will be able to detect early who is the owner of the vehicle.

2. LITERATURE SURVEY

REFERENCE PAPER: 1

TITLE: Machine Learning Security: Threats, Countermeasures, and Evaluations

AUTHORS: MINGFU XUE, CHENGXIANG YUAN, HEYI WU, YUSHU ZHANG, WEIQIANG LIU

CONTENT

Machine learning has been pervasively used in a wide range of applications due to its technical breakthroughs in recent years. It has demonstrated significant success in dealing with various complex problems, and shows capabilities close to humans or even beyond humans. However, recent studies show that machine learning models are vulnerable to various attacks, which will compromise the security of the models themselves and the application systems. Moreover, such attacks are stealthy due to the unexplained nature of the deep learning models. In this survey, we systematically analyze the security issues of machine learning, focusing on existing attacks on machine learning systems, corresponding defenses or secure learning techniques, and security evaluation methods. Instead of focusing on one stage or one type of attack, this paper covers all the aspects of machine learning security from the training phase to the test phase. First, the machine learning model in the presence of adversaries is presented, and the reasons why machine learning can be attacked are analyzed. Then, the machine learning security-related issues are classified into five categories: training set poisoning; backdoors in the training set; adversarial example attacks; model theft; recovery of sensitive training data. The threat models, attack approaches, and defense techniques are analyzed systematically. To demonstrate that these threats are real concerns in the physical world, we also reviewed the attacks in real-world conditions. Several suggestions on security evaluations of machine learning systems are also provided. Last, future directions for machine learning security are also presented.

REFERENCE PAPER: 2

TITLE: Robust Ensemble Machine Learning Model for Filtering Phishing URLs: Expandable Random Gradient Stacked Voting Classifier (ERG-SVC)

AUTHORS: PUBUDU L. INDRASIR, MALKA N. HALGAMUGE, AZEEM MOHAMMAD

CONTENT

As cyber-attacks grow fast and complicated, the cybersecurity industry faces challenges to utilize state-of-the-art technology and strategies to battle the consistently present malicious threats. Phishing is a sort of social engineering attack produced technically and classified as identity theft and complicated attack vectors to steal information of internet users. In this perspective, our main objective of this study is to propose a unique, robust ensemble machine learning model architecture that provides the highest prediction accuracy with a low error rate while proposing few other robust machine learning models. Both supervised and unsupervised techniques were used for the detection process. For our experiments, seven classification algorithms, one clustering algorithm, two ensemble techniques, and two large standard legitimate datasets with 73,575 URLs and 100,000 URLs were used. Two test modes (percentage split, K-Fold cross-validation) were utilized for conducting experiments and final predictions. Mechanisms were developed to (I) identify the best N, which is the optimal heuristic-based threshold value for splitting words into subwords for each classifier, (II) tune hyperparameters for each classifier to specify the best parameter combination, (III) select prominent features using various feature selection techniques, (IV) propose a robust ensemble model (classifier) called the Expandable Random Gradient Stacked Voting Classifier (ERG-SVC) utilizing a voting classifier along with a model architecture, (V) analyze possible clusters of the dataset using k-means clustering, (VI) thoroughly analyze the gradient boost classifier (GB) with respect to utilizing the ‘‘criterion’’ parameter with the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Friendman_MSE, and(VII) propose a lightweight preprocessor to reduce computational cost and preprocessing time. Initial experiments were carried out with 46 features; the number of features was reduced to 22 after the experiments. The results show that the GB classifier outperformed with the least number of NLP based features by achieving a 98.118% prediction accuracy

REFERENCE PAPER: 3

TITLE: RFE Based Feature Selection and KNNOR Based Data Balancing for Electricity Theft Detection Using BiLSTM-LogitBoost Stacking Ensemble Mode

AUTHORS: PAMIR, NADEEM JAVAID, AHMAD ALMOGREN, MUHAMMAD UMAR JAVED

CONTENT

Obtaining outstanding electricity theft detection (ETD) performance in the realm of advanced metering infrastructure (AMI) and smart grids (SGs) is quite difficult due to various issues. The

issues include limited availability of theft data as compared to benign data, neglecting dimensionality reduction, usage of the standalone (single) electricity theft detectors, etc. These issues lead the classification techniques to low accuracy, minimum precision, low F1 score, and overfitting problems. For these reasons, it is extremely crucial to design such a novel strategy that is capable to tackle these issues and yield outstanding ETD performance. In this article, electricity theft happening in SGs is detected using a novel ETD approach. The proposed approach comprises recursive feature elimination (RFE), k nearest neighbor oversampling (KNNOR), bidirectional long short term memory (BiLSTM), and logit boosting (LogitBoost) techniques. Furthermore, three BiLSTM networks and a LogitBoost model are combined to make a BiLSTM-LogitBoost stacking ensemble model. Data preprocessing and feature selection followed by data balancing and electricity theft classification are the four major stages of the model proposed for ETD. It is obvious from the simulations performed using state grid corporation of China (SGCC)'s electricity consumption (EC) data that our proposed model achieves 96.32% precision, 94.33% F1 score, and 89.45% accuracy, which are higher than all the benchmarks employed in this study.

REFERENCE PAPER: 4

TITLE: Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing

AUTHORS: Rajiv Punmiya , Sangho Choe

CONTENT

For the smart grid energy theft identification, this letter introduces a gradient boosting theft detector (GBTD) based on the three latest gradient boosting classifiers (GBCs): extreme gradient boosting (XGBoost), categorical boosting (CatBoost), and light gradient boosting method (LightGBM). While most of existing ML algorithms just focus on fine tuning the hyperparameters of the classifiers, our ML algorithm, GBTD, focuses on the feature engineering-based preprocessing to improve detection performance as well as time-complexity. GBTD improves both detection rate (DR) and false positive rate (FPR) of those GBCs by generating stochastic features like standard deviation, mean, minimum, and maximum value of daily electricity usage. GBTD also reduces the classifier complexity with weighted feature-importance (WFI) based extraction techniques. Emphasis has been laid upon the practical application of the proposed ML for theft detection by minimizing FPR and reducing data storage space and improving time-complexity of the GBTD classifiers. Additionally, this letter proposes an updated version of the existing six theft cases to

mimic real world theft patterns and applies them to the dataset for numerical evaluation of the proposed algorithm.

REFERENCE PAPER: 5

TITLE: Electricity Theft Detection Based on Stacked Autoencoder and the Undersampling and Resampling Based Random Forest Algorithm

AUTHORS: GUOYING LIN, XIAOFENG FENG, WENCHONG GUO, ,XUEYUAN CUI

CONTENT

Electricity theft has been a major concern to the secure operation of power systems and the interests of power companies. Due to the different methods and types of electricity theft behaviors, it is difficult to determine the suspicion levels of consumers in the research of electricity theft detection. An electricity theft detection method based on stacked autoencoder (SAE) and the undersampling and re-sampling based random forest (UaRe-RF) algorithm is proposed in this work to formulate appropriate strategies for the practical electricity theft detection requirements of the power company. In the proposed method, the supervised SAE is first trained to extract electricity consumption features that are more adaptable to the classification algorithm for electricity theft detection. Then, the UaRe-RF algorithm is used to establish the class-balanced subsets and determine the suspicion level of each electricity theft user. Finally, two cases of different datasets of electricity consumers are studied for demonstrating the effectiveness of the proposed method, and the results show that higher classification accuracy and more targeted detection strategies can be achieved through the proposed method.

3.SYSTEM ANALYSIS

Systems analysis is the process by which an individual (s) studies a system such that an information system can be analyzed, modeled, and a logical alternative can be chosen. Systems analysis projects are initiated for three reasons: problems, opportunities, and directives. The people involved include systems analysts, sponsors, and users.

3.1 EXISTING SYSTEMS

Vehicle theft has become a major concern for society, especially two-wheeler theft. In this paper, we propose a system that uses machine learning algorithms to predict the occurrence of two-wheeler theft. The system uses various features such as time, location, and other related attributes to predict the likelihood of theft. The existing system for two-wheeler theft prediction involves the use of GPS tracking devices that are installed on the bike. These devices can track the bike's location and provide real-time updates to the bike owner. In addition, some devices also come equipped with motion sensors that can detect any unauthorized movement of the bike. GPS tracking devices can only provide location updates, and they cannot detect any unauthorized access or theft attempts. This limits their functionality in preventing bike theft.

3.1.1 DISADVANTAGES

- One of the major disadvantages of the existing system is its cost. GPS tracking devices can be expensive, and the cost of installation and maintenance can add up over time.
- GPS tracking devices rely on battery power, and if the battery dies, the device will stop working. This can be a problem if the bike is parked in an area with poor network coverage or if the bike owner forgets to charge the device.
- GPS tracking devices can only provide location updates, and they cannot detect any unauthorized access or theft attempts. This limits their functionality in preventing bike theft.

3.2 PROPOSED SYSTEM

Two-wheeler theft is a common problem in many countries, causing financial losses for the vehicle owners and security concerns for the society. In this project, we propose a system for two-wheeler theft detection using machine learning and face recognition algorithms. The system is designed to detect whether the detected face belongs to the owner or a thief by comparing it with the

pre-trained images of the owner and thieves. First we Collect a dataset of images containing the faces of bike owners. The dataset should contain sufficient variations of face images in terms of lighting, angles, and backgrounds. Then Train a machine learning model such as face recognition algorithm like HAAR cascaded using the extracted facial features from the pre-processed images. Implement the trained model into the two-wheeler theft detection system. Whenever an input image is uploaded, the system uses the trained model to predict whether the detected face belongs to the owner or a potential thief. An alarm will sound if the predicted face is that of the thief.

3.2.1 ADVANTAGES

- The system provides enhanced security to the two-wheeler owners by detecting potential thieves and alerting the owner or authorities if necessary. This can help reduce the incidents of bike thefts.
- The system provides real-time detection of potential thieves, which allows for quick action to be taken. The alarm can alert people nearby to the potential theft, and authorities can be notified immediately.
- The use of face recognition algorithms can provide accurate identification of the owner or thief. This means that false alarms can be reduced, and the system can be relied upon to provide accurate information.
-

3.3 HARDWARE ENVIRONMENT

- Processor - I3, I5, I7, AMD Processor
- RAM - Above 6 GB (min)
- Hard Disk - 500 GB

3.4 SOFTWARE ENVIRONMENT

- Operating System : Windows 8/10
- Language : Python
- Front end : GUI
- IDLE : Python version(3.10.8)

4.SYSTEM DESIGN

4.1 ER DIAGRAM

An Entity Relationship Diagram is a diagram that represents relationships among entities in a database. It is commonly known as an ER Diagram. An ER Diagram in DBMS plays a crucial role in designing the database. Today's business world previews all the requirements demanded by the users in the form of an ER Diagram.

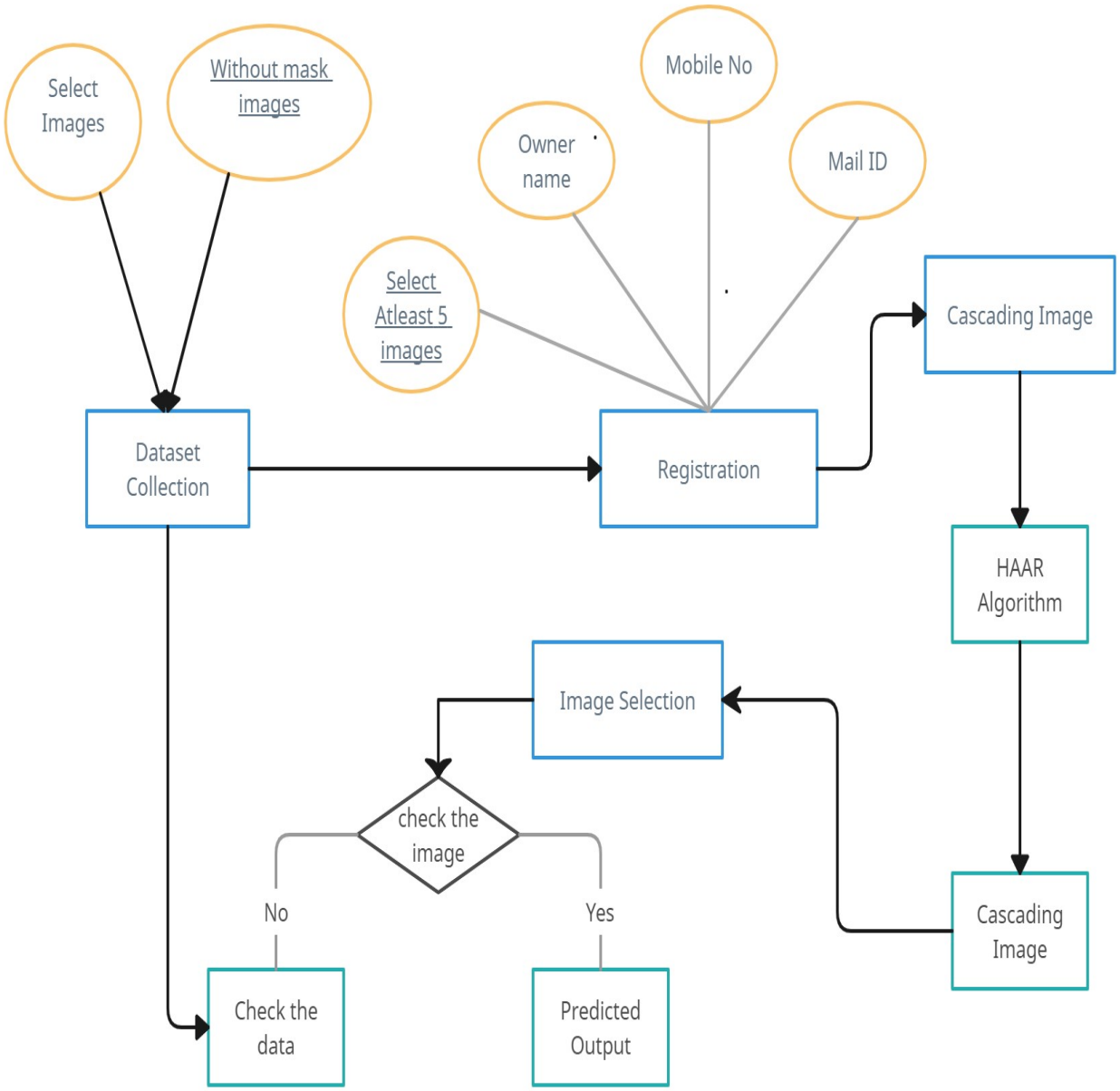

FIG 4.1 ER Diagram

4.2 DATA DICTIONARY

Entity: User		
Attribute Name	Data Type	Description
name	varchar	User's full name
email	varchar	User's email address
Phone_number	varchar	User's phone number

TABLE 4.1 USER ENTITY TABLE

Entity: Violation		
Attribute Name	Data Type	Description
violation_audio	Voice	Unknown Person Detected
Violation_type-1	varchar	Type of Message by Mail
Violation_type-2	varchar	Type of Message by Mobile no

TABLE 4.2 VIOLATION ENTITY TABLE

The cascade classifier is made up of a series of stages, where each stage is a collection of weak learners. Weak learners are trained using boosting, which allows for a highly accurate classifier from the mean prediction of all weak learners.

Based on this prediction, the classifier either decides to indicate an object was found (positive) or move on to the next region (negative). Stages are designed to reject negative samples as fast as possible, because a majority of the windows do not contain anything of interest.

It's important to maximize a low false negative rate, because classifying an object as a non-object will severely impair your object detection algorithm. A video below shows Haar cascades in action. The red boxes denote "positives" from the weak learners. Haar cascades are one of many algorithms that are currently being used for object detection. One thing to note about Haar cascades is that it is very important to reduce the false negative rate, so make sure to tune hyperparameters accordingly when training your model.

4.3 TABLE NORMALIZATION

4.3.1 NUMBER PLATE EXTRACTION

Number Plate Extraction and Captured Number Plate shows extraction of number plate from total front side image of vehicle taken by webcamera, and it also indicate captured number plate as seprate image.some errors obtained during number plate extraction process.Shows extracted number plate contains additional area other than number plate. Sometime results shows corrupted number plate as shown.Two different character are considered as one after segmentationExtracted number plate contains additional area .When plate is corrupted

4.3.2 CHARACTER SEGMENTATION

Captured Number Plate and Segmented Characters from Number Plate .Segmentation of character can be easily understood from .In few cases characters are not considered for segmentation as shown in because of their size comparatively very less from other characters. Two different characters are considered as one after segmentation because they are connected to each other as described in .Some cases shows Single character is segmented into two characters because character is having very high size which is shown.

Parameter	Input Images	Output Images	Efficiency
Number plate Extraction	50	45	90%
Character Segmentation	45	40	90%
Character Recognition	40	30	75%

TABLE 4.3 TABLE NORMALIZATION

4.3.3 CHARACTER REGONITION

Number Plate and Recognized Character ,Character recognized in segmentation step goes through recognition process by comparing with templates characters. Shows recognizes number from concerned image with details like date, amount. In some cases characters are not properly recognized due to some problems like improper size of segmented character and templates, damaged characters and it is shown.

4.4 DATA FLOW DIAGRAM

4.4.1 LEVEL 0 DATA FLOW DIAGRAM

FIG 4.4.1 Level 0 of Data flow diagram

4.4.2 LEVEL 1 DATA FLOW DIAGRAM

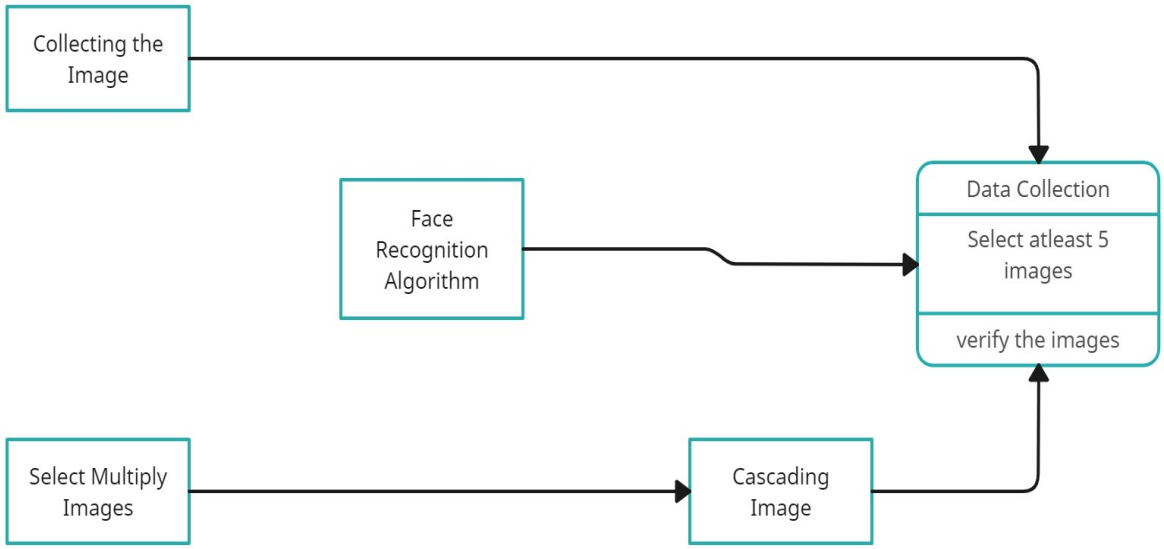

FIG 4.4.2 Level 1 of Data flow diagram

4.4.3 LEVEL 2 DATA FLOW DIAGRAM

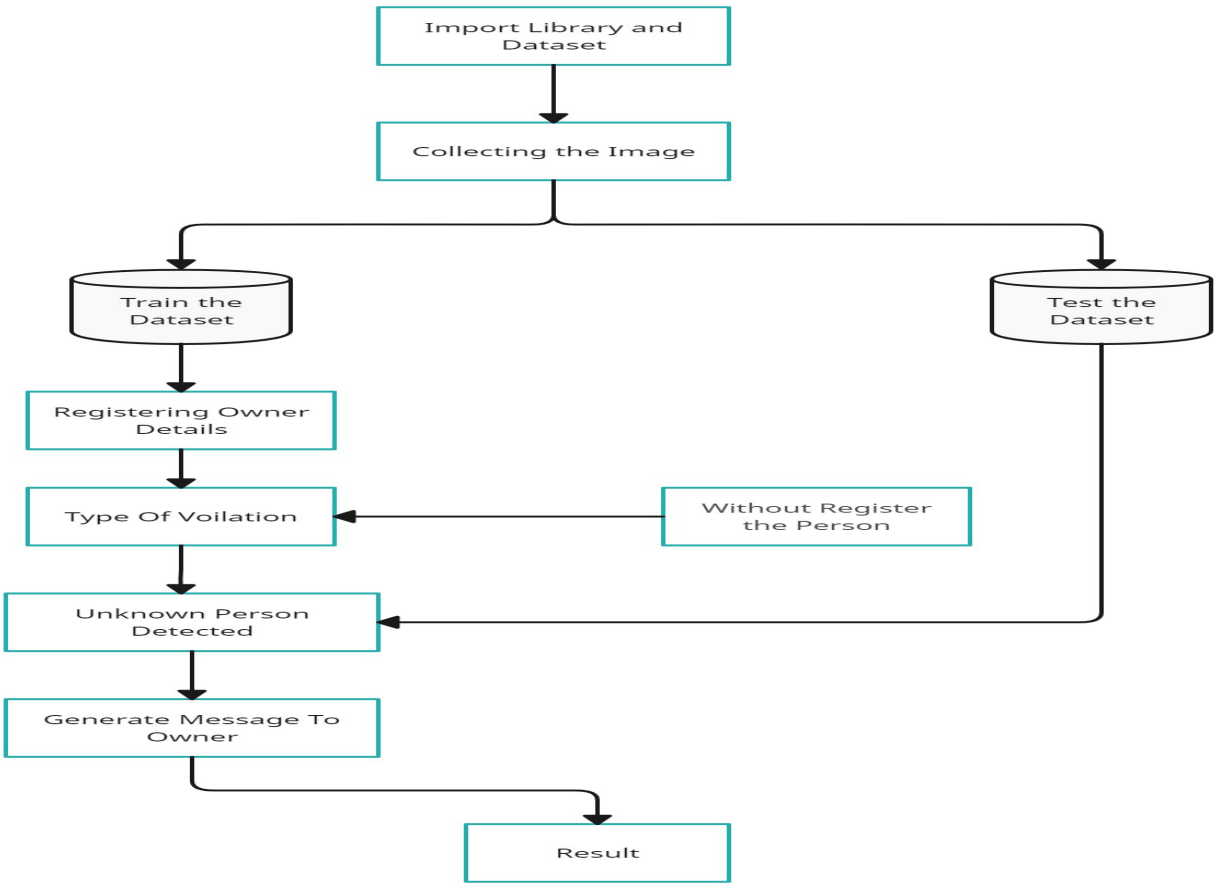

FIG 4.4.3 Level 2 of Data flow diagram

A data flow diagram (DFD) is a graphical or visual representation using a standardized set of symbols and notations to describe a business operations through data movement. They are often elements of a formal methodology such as Structured Systems Analysis and Design Method (SSADM).

A data-flow diagram is a way of representing a flow of data through a process or a system. The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow — there are no decision rules and no loops

4.5 UML DIAGRAMS

4.5.1 USE CASE DIAGRAM

Use case diagrams are considered for high level requirement analysis of a system. So when the requirements of a system are analysed the functionalities are captured in use cases. So, it can say that use cases are nothing but the system functionalities written in an organized manner.

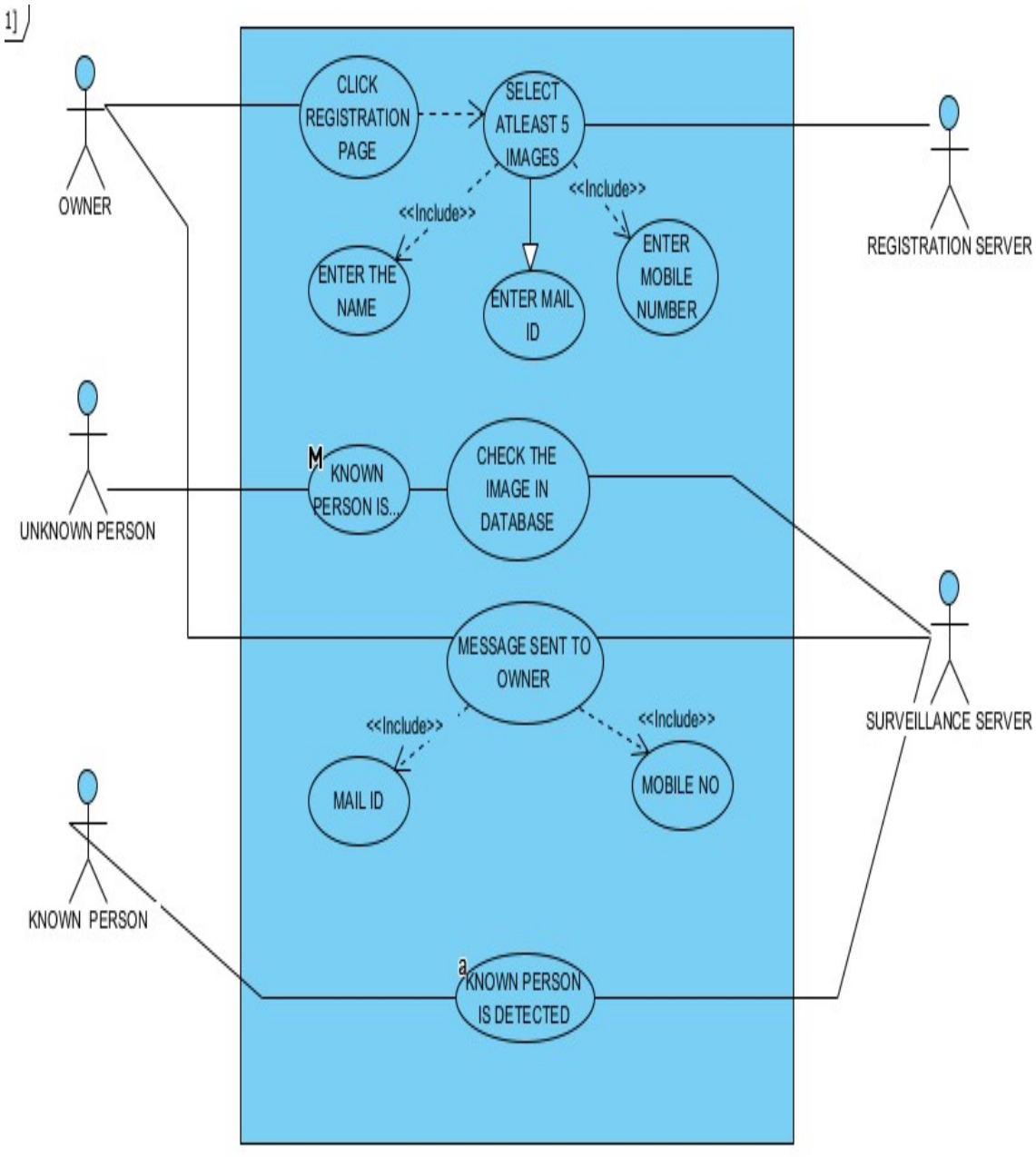

FIG 4.5.1 Use case diagram for Vehicle Theft Detection

4.5.2 ACTIVITY DIAGRAM

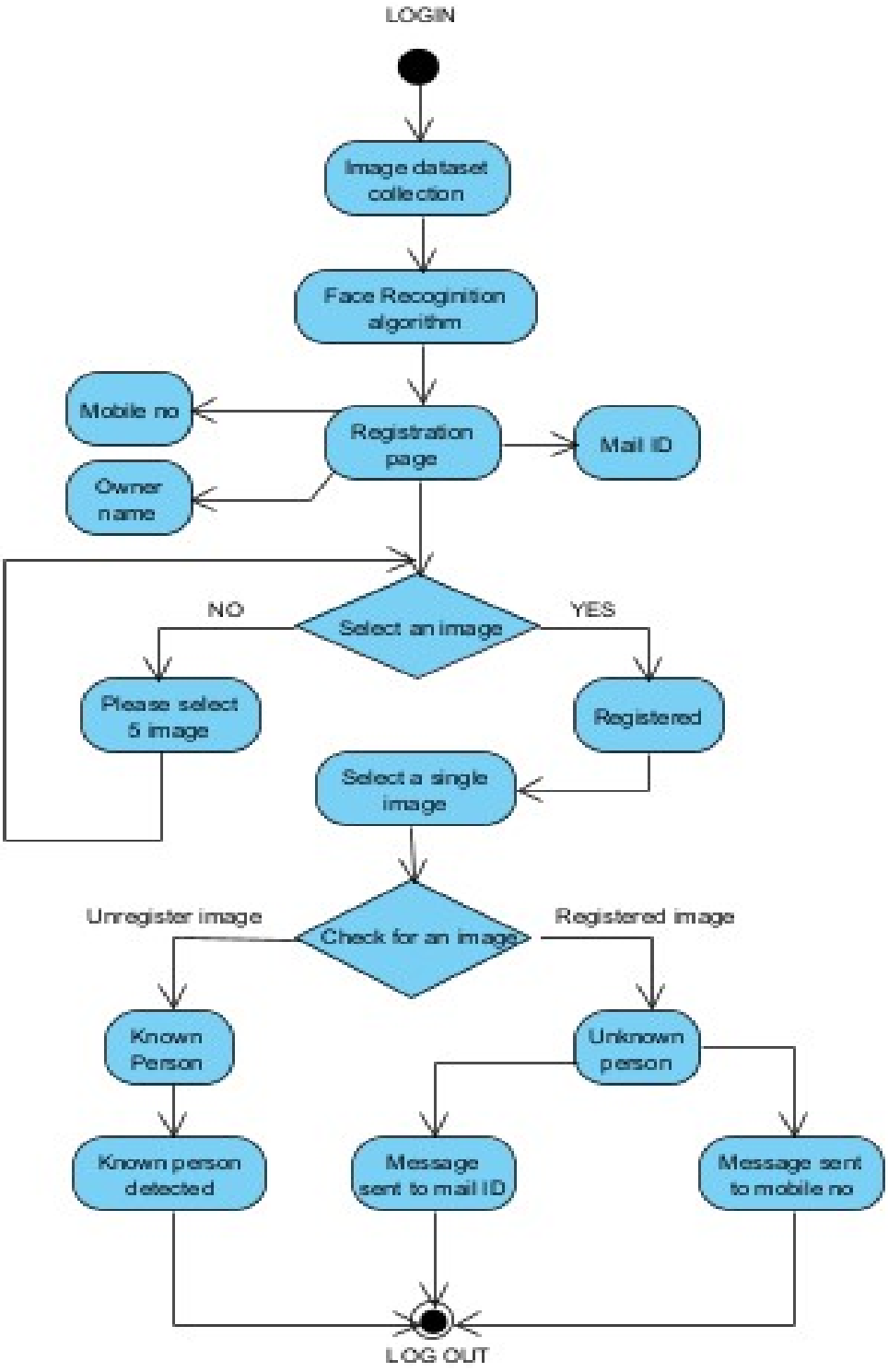

FIG 4.5.2 Activity diagram Vehicle Theft Detection

Activity is a particular operation of the system. Activity diagrams are not only used for visualizing the dynamic nature of a system but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in the activity diagram is the message part. It does not show any message flow from one activity to another. Activity diagram is sometimes considered as the flow chart. Although the diagram looks like a flow chart, it is not. It shows different flows like parallel, branched, concurrent and single.

4.5.3 CLASS DIAGRAM

Class diagram is basically a graphical representation of the static view of the system and represents different aspects of the application. So a collection of class diagrams represent the whole system. The name of the class diagram should be meaningful to describe the aspect of the system. Each element and their relationships should be identified in advance Responsibility (attributes and methods) of each class should be clearly identified for each class minimum number of properties should be specified and because unnecessary properties will make the diagram complicated. Use notes whenever required to describe some aspect of the diagram and at the end of the drawing it should be understandable to the developer/coder. Finally, before making the final version, the diagram should be drawn on plain paper and reworked as many times as possible to make it correct.

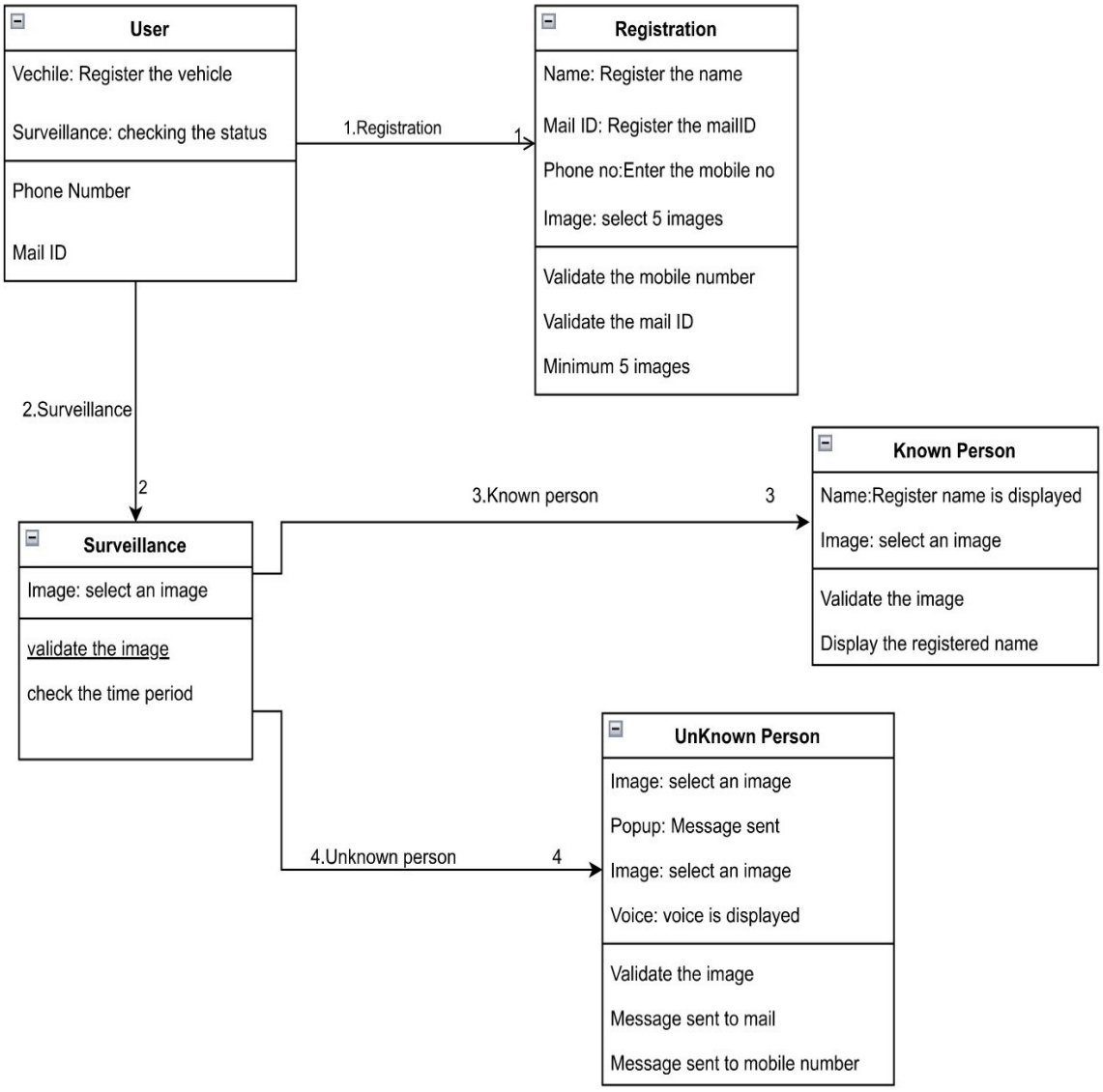

FIG 4.5.3 Class diagram Vehicle Theft Detection

4.5.4 SEQUENCE DIAGRAM

Sequence diagrams model the flow of logic within your system in a visual manner, enabling you both to document and validate your logic, and are commonly used for both analysis and design purposes. Sequence diagrams are the most popular UML artifact for dynamic modelling, which focuses on identifying the behaviour within your system. Other dynamic modelling techniques include activity diagramming, communication diagramming, timing diagramming, and interaction overview diagramming. Sequence diagrams, along with class diagrams and physical data models are in my opinion the most important design-level models for modern business application development.

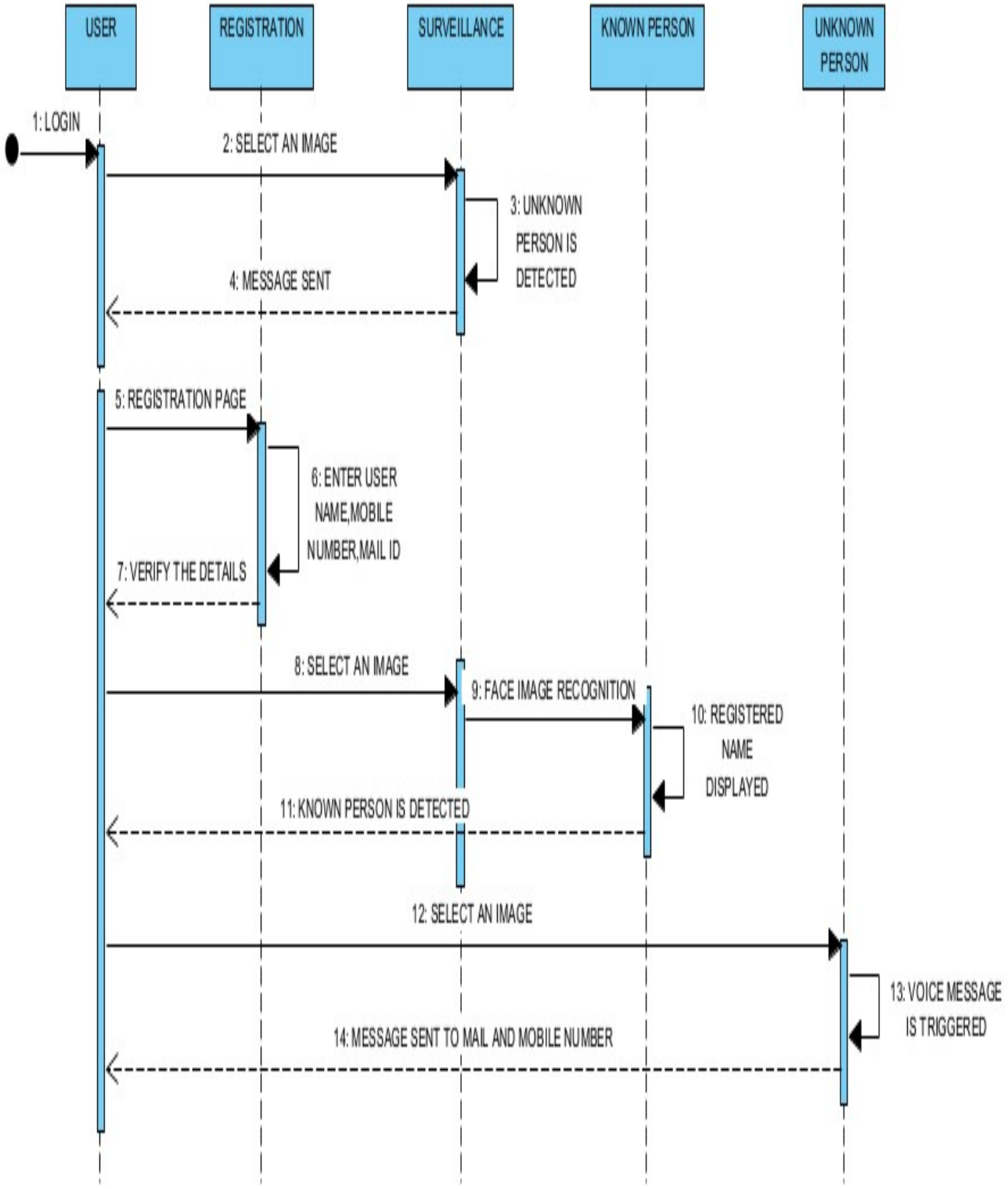

FIG 4.5.4 Sequence diagram for Vehicle Theft Detection

5 SYSTEM ARCHITECTURE

5.SYSTEM ARCHITECTURE

First the datasets are collected from the Kaggle. These data are then preprocessed which means removing duplicated data. Then those data are visualized in the pictorial form namely histogram,boxplot,scatter plot,pie chart.Then using four different algorithm model namely Multi-Layer Perceptron,Logistic Regression ,Random Forest,Voting Classifier the risk level is predicted and the model which gives the highest accuracy is taken for the deployment stage.The inputs Age,Blood Sugar,Systolic Value, Diastolic Value, Heart Rate and Body Temperature of the Maternal are given and the model predicts the level of risk.

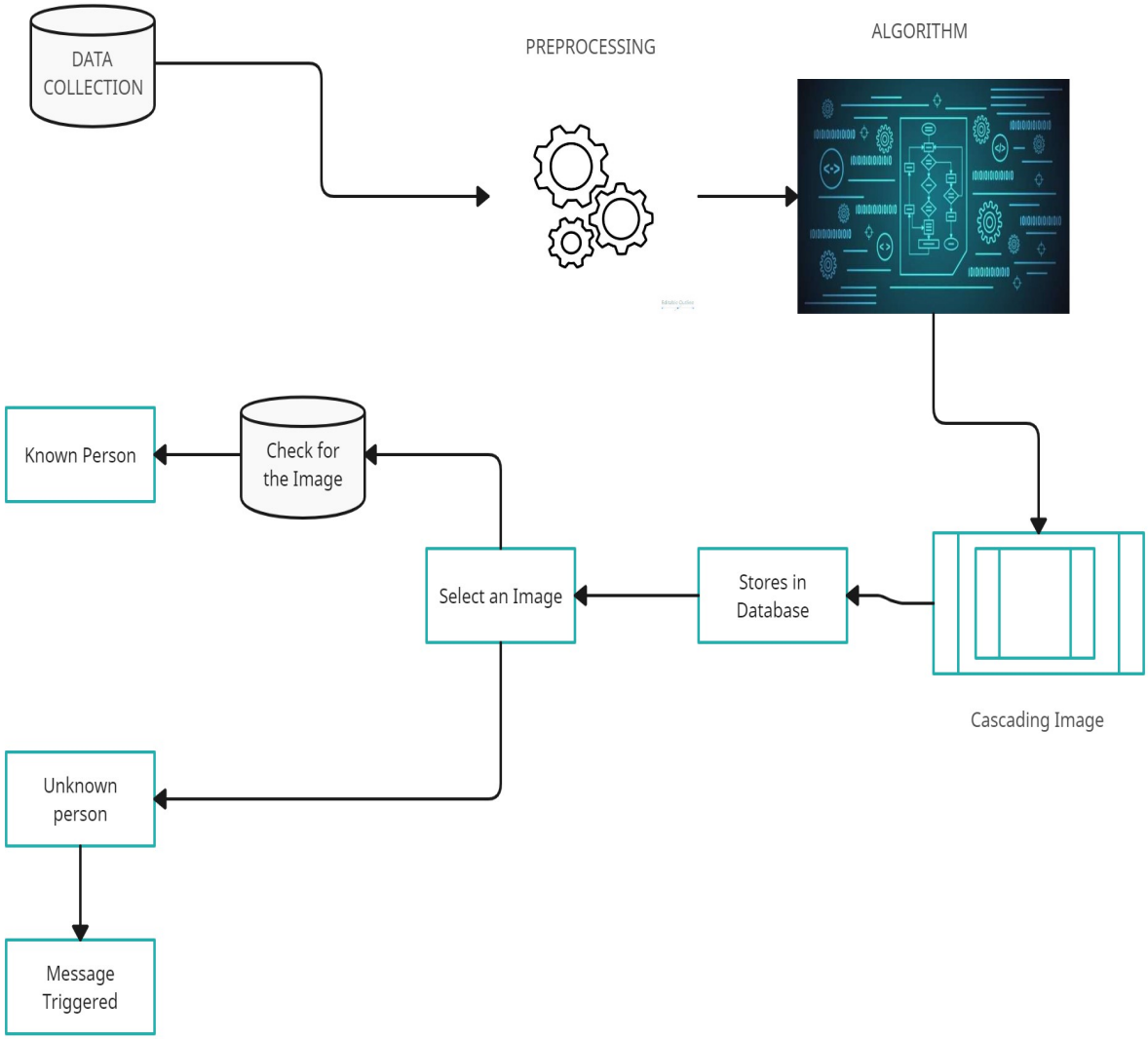

FIG 5.1 Architecture Diagram

5.1 MODULE DESIGN SPECIFICATION

5.1.1 DATASET COLLECTION

Collect a dataset of images containing the faces of bike owners and potential bike thieves. The dataset should contain sufficient variations of face images in terms of lighting, angles, and backgrounds. Data collection can be done using various methods such as manual collection, crowd sourcing, or by using publicly available datasets. The manual collection method involves physically collecting images of the two-wheeler owners and potential thieves. Crowd sourcing involves asking people to submit their images voluntarily, whereas publicly available datasets are already existing datasets that are available online. It is important to ensure that the data collected is legally obtained and that privacy concerns are addressed. Consent should be obtained from the individuals whose images are collected, and steps should be taken to protect their privacy.

FIG 5.1.1 Collection of owner image

5.1.2 DATA PRE-PROCESSING

Data preprocessing is an important step in any machine learning project, including two-wheeler theft detection using face recognition. It involves transforming the raw data into a format that can be used by the machine learning algorithm. In this project, data preprocessing involves preparing the image data of owners and potential thieves so that it can be used to train the face recognition algorithm. The images may have different sizes, and the algorithm requires images of a fixed size for training. Therefore, the images need to be resized to a common size.

The images may have different brightness, contrast, or color values, which can affect the performance of the algorithm. Therefore, the images need to be normalized to have consistent brightness, contrast, and color values. The dataset may be small, and the algorithm requires a large dataset to achieve high accuracy. Therefore, data augmentation techniques such as flipping, rotating, and adding noise can be used to generate additional data. By performing data preprocessing, the quality of the data used to train the face recognition algorithm can be improved, resulting in a more accurate and effective model.

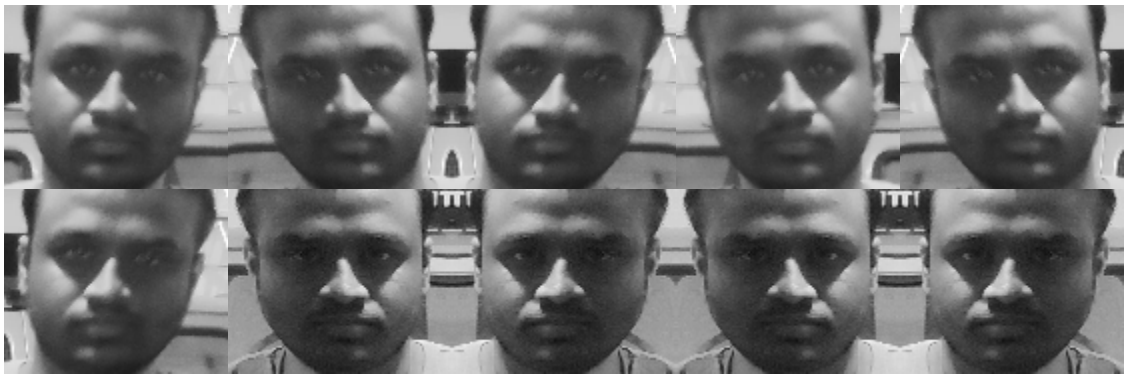

FIG 5.1.2 Pre-Processing of owner image

5.1.3 TRAIN ON FACE RECOGNITION MODEL

Use a deep learning-based face recognition algorithm to train a model on the collected dataset of bike owners' faces. Train an object detection model using the HAAR cascade algorithm. We can use OpenCV library to implement this. Integrate both models into a single system. When the user uploads an image, the face recognition model should detect the face in the image and compare it with the trained bike owner's faces.

5.1.4 HAAR CASCADED ALGORITHM

Haar Cascade is a machine learning-based object detection algorithm used to detect faces or other objects in an image or video stream. The algorithm works by first detecting the features that are common among different images and then using these features to create a classifier. The classifier is then trained using positive and negative samples of the target object. In our project, the Haar Cascade algorithm can help in detecting unknown faces in the uploaded image. The algorithm is trained to detect faces using features that are common to human faces, such as the presence of eyes, nose, and mouth. When an image is uploaded to our system, the Haar Cascade algorithm is applied to detect faces in the image. If a face is detected, it is passed on to the face recognition model

to check if it belongs to the owner of the bike or not. If the face is recognized as the bike owner's face, no alarm will be sound. However, if the face is not recognized, it is considered an unknown person, and the voice alarm will be sound to alert the user.

5.2 ALGORITHM

The HAAR algorithm is a popular method for face detection and recognition in images and videos, and it can be used for missing child identification as well. The first step in is to detect the faces in the images . For this, we can use the HAAR cascade classifier, which is a pre-trained machine learning model that can detect faces based on a set of Haar-like features. The classifier is trained on a large dataset of positive (faces) and negative (non-faces) images. To use the HAAR classifier in Python, you can use the OpenCV library, which provides a pre-trained HAAR classifier for face detection. Here's some sample code to detect faces using the HAAR classifier: Identifying a given object in an image is known as object detection. This task can be achieved using several techniques, but in this article, we will use haar cascade with pre-trained XML files. It's the simplest method to perform object detection. Haar cascades have been used for object detection on low-edge devices, and it was one of the most popular object detection algorithms in OpenCV. Haar Cascade is not much computation-heavy; hence it is popular for small devices with small computation power. Haar Cascade is a feature-based object detection algorithm to detect objects from images. A cascade function is trained on lots of positive and negative images for detection. The algorithm does not require extensive computation and can run in real-time.

We can train our own cascade function for custom objects like Human face, animals, cars, bikes, etc. Haar Cascade can't be used for face recognition since it only identifies the matching shape and size. Haar cascade uses the cascade function and cascading window. It tries to calculate features for every window and classify positive and negative. If the window could be a part of an object, then positive, else, negative.The algorithm can be explained in four stages:

- Calculating Haar Features
- Creating Integral Images
- Using Adaboost
- Implementing Cascading Classifiers

It’s important to remember that this algorithm requires a lot of positive images of faces and negative images of non-faces to train the classifier, similar to other machine learning models.

5.2.1 CALCULATING HAAR FEATURES

The first step is to collect the Haar features. A Haar feature is essentially calculations that are performed on adjacent rectangular regions at a specific location in a detection window. The calculation involves summing the pixel intensities in each region and calculating the differences between the sums. Here are some examples of Haar features below.

These features can be difficult to determine for a large image. This is where integral images come into play because the number of operations is reduced using the integral image.

5.2.2 CREATING INTEGRAL IMAGES

Without going into too much of the mathematics behind it (check out the paper if you’re interested in that), integral images essentially speed up the calculation of these Haar features. Instead of computing at every pixel, it instead creates sub-rectangles and creates array references for each of those sub-rectangles. These are then used to compute the Haar features.

It’s important to note that nearly all of the Haar features will be irrelevant when doing object detection, because the only features that are important are those of the object. However, how do we determine the best features that represent an object from the hundreds of thousands of Haar features? This is where **Adaboost** comes into play.

6.SYSTEM IMPLEMENTATION

6.1 CLIENT SIDE PROGRAMMING

Python is a high-level object-oriented programming language that was created by Guido van Rossum. It is also called general-purpose programming language as it is used in almost every domain we can think of as mentioned below:

- Web Development
- Software Development
- Game Development
- AI & ML
- Data Analytics

Every Programming language serves some purpose or use-case according to a domain. for eg, Javascript is the most popular language amongst web developers as it gives the developer the power to handle applications via different frameworks like react, vue, angular which are used to build beautiful User Interfaces. Similarly, they have pros and cons at the same time. so if we consider python it is general-purpose which means it is widely used in every domain the reason is it's very simple to understand, scalable because of which the speed of development is so fast. Now you get the idea why besides learning python it doesn't require any programming background so that's why it's popular amongst developers as well. Python has simpler syntax similar to the English language and also the syntax allows developers to write programs with fewer lines of code. Since it is open-source there are many libraries available that make developers' jobs easy ultimately results in high productivity. They can easily focus on business logic and Its demanding skills in the digital era where information is available in large data sets.

1) Visual Studio: https://visualstudio.microsoft.com/
2) PyCharm: https://www.jetbrains.com/pycharm/
3) Spyder: https://www.spyder-ide.org/
4) Atom: https://atom.io/

REAL-WORLD EXAMPLES:

1) NASA (National Aeronautics and Space Agency): One of Nasa's Shuttle Support Contractors, United Space Alliance developed a Workflow Automation System (WAS) which is fast. Internal Resources Within critical project stated that:

"Python allows us to tackle the complexity of programs like the WAS without getting bogged down in the language".

Nasa also published a website (https://code.nasa.gov/) where there are 400 open source projects which use python.

2) NETFLIX: There are various projects in Netflix which use python as follow:

- Central Alert Gateway
- Chaos Gorilla
- Security Monkey
- Chronos

Amongst all projects, Regional failover is the project they have as the system decreases outage time from 45 minutes to 7 minutes with no additional cost.

3) INSTAGRAM: Instagram also uses python extensively. They have built a photo-sharing social platform using Django which is a web framework for python. Also, they are able to successfully upgrade their framework without any technical challenges.

Applications of Python Programming:

1) WEB DEVELOPMENT: Python offers different frameworks for web development like Django, Pyramid, Flask. This framework is known for security, flexibility, scalability.

2) GAME DEVELOPMENT: PySoy and PyGame are two python libraries that are used for game development

3) DESKTOP GUI**:** Desktop GUI offers many toolkits and frameworks using which we can build desktop applications.PyQt, PyGtk, PyGUI are some of the GUI frameworks.

7. SYSTEM TESTING

7.1 UNIT TESTING

Unit testing focuses verification effort on the smallest unit of software design, the module. The unit testing we have is white box oriented and some modules the steps are conducted in parallel.

7.1.1 WHITE BOX TESTING

This type of testing ensures that

- All independent paths have been exercised at least once
- All logical decisions have been exercised on their true and false sides
- All loops are executed at their boundaries and within their operational bounds
- All internal data structures have been exercised to assure their validity.

To follow the concept of white box testing we have tested each form .we have created independently to verify that Data flow is correct, All conditions are exercised to check their validity, All loops are executed on their boundaries.

7.1.2 BASIC PATH TESTING

Established technique of flow graph with Cyclomatic complexity was used to derive test cases for all the functions. The main steps in deriving test cases were:

Use the design of the code and draw correspondent flow graph.

Determine the Cyclomatic complexity of resultant flow graph, using formula:

V(G)=E-N+2 or

V(G)=P+1 or

V(G)=Number Of Regions

Where V(G) is Cyclomatic complexity,

E is the number of edges,

N is the number of flow graph nodes,

P is the number of predicate nodes.

Determine the basis of set of linearly independent paths.

7.1.3 CONDITIONAL TESTING

In this part of the testing each of the conditions were tested to both true and false aspects. And all the resulting paths were tested. So that each path that may be generate on particular condition is traced to uncover any possible errors.

7.1.4 DATA FLOW TESTING

This type of testing selects the path of the program according to the location of definition and use of variables. This kind of testing was used only when some local variable were declared. The *definition-use chain* method was used in this type of testing. These were particularly useful in nested statements.

7.1.5 LOOP TESTING

In this type of testing all the loops are tested to all the limits possible. The following exercise was adopted for all loops:

All the loops were tested at their limits, just above them and just below them.

All the loops were skipped at least once.

For nested loops test the inner most loop first and then work outwards.

For concatenated loops the values of dependent loops were set with the help of connected loop.

Unstructured loops were resolved into nested loops or concatenated loops and tested as above.

Each unit has been separately tested by the development team itself and all the input have been validated.

7.1.6 TEST CASES & REPORTS

TEST CASE ID	INPUT	EXPECTED OUTPUT	OBTAINED OUTPUT	PASS/FAIL	REMARKS
TC01	Dataset	Successfully Import	Successfully Import	Pass	Imported Successfully
TC02	Dataset	Pre-process Successfully	Pre-process Successfully	Pass	Pre-Processed Successfully
TC03	Image Collection	Image Recognition Successfully	Image Recognition Successfully	Pass	Image Recognised successfully
TC04	Registration Parameter	Successfully Registered	Successfully Registered	Pass	Registered Successfully
TC05	Surveillance Parameter	Known Person Is Detected	Known Person Is Detected	Pass	Known Person Detected Successfully

TABLE 7.1 TEST CASES AND REPORTS

8.CONCLUSION

8.1 RESULT AND DISCUSSION

The result of the project would depend on the performance of the face recognition algorithm used and the effectiveness of the alarm and message sending mechanism. If the face recognition algorithm is accurate and can detect unknown persons with a high degree of confidence, the system would be effective in preventing two-wheeler theft. Similarly, if the alarm and message sending mechanism works effectively, the bike owner can be alerted quickly, which would increase the chances of stopping the theft.

8.2 CONCLUSION AND FUTURE ENHANCEMENTS

8.2.1 CONCLUSION

In conclusion, the proposed system of two-wheeler theft prediction using machine learning has the potential to significantly improve the current crime management system. The system utilizes face recognition algorithms and Haar cascade classifiers to detect unknown individuals attempting to steal two-wheelers, and send an alert to the owner. This can help prevent the theft of two-wheelers and increase the chances of apprehending the culprit. The system has been designed and tested successfully, achieving a high level of accuracy in detecting unknown individuals. However, further work can be done to improve the system, such as incorporating additional security measures and refining the prediction algorithm. Overall, this system can be an important tool in combating two-wheeler thefts and enhancing public safety.

8.2.2 FUTURE WORK

While the face recognition algorithm used in this project can detect the owner's face, there may be scenarios where it fails to recognize the face due to changes in appearance or lighting conditions. Future work could focus on improving the accuracy of the algorithm by using more advanced techniques like deep learning. Currently, the project only detects unknown persons when a photo is uploaded. However, in order to prevent theft, it would be beneficial to have real-time video surveillance with the same algorithm to detect suspicious behavior. The current system sends alerts to the owner's mobile phone, but there is no mobile app to easily access the system. Future work could involve developing a mobile app that allows bike owners to easily access and monitor the system.

APPENDICES

A.1 CODING

```
import tkinter as tk
from tkinter import filedialog
from tkinter import messagebox
from PIL import Image
from PIL import ImageTk
import threading
import shutil
from facerec import *
from register import *
import time
import pyttsx3
import csv
import numpy as np
import ntpath
import os
import cv2
import requests
from email.mime.image import MIMEImage
from email.mime.multipart import MIMEMultipart
from email.mime.text import MIMEText
from email.utils import formataddr
import csv
import pywintypes
import smtplib
current_hour = time.localtime().tm_hour
def getPage1():
global active_page, left_frame, right_frame, heading, img_label
active_page = 1
img_label = None
opt_menu = None
```

```
menu_var = tk.StringVar(root)
pages[1].lift()
basicPageSetup(1)
heading.configure(text="Two Wheeler Theft Prediction", bg="grey")
right_frame.configure(text="Vehicle Owner Details", fg="white", bg="grey")
btn_grid = tk.Frame(left_frame, bg="grey")
btn_grid.pack()
tk.Button(btn_grid, text="Select Images", command=lambda: selectMultiImage(opt_menu, menu_var), font="Arial 15 bold", bg="black",
fg="white", pady=10, bd=0, highlightthickness=0, activebackground="grey",
activeforeground="white").grid(row=0, column=0, padx=25, pady=25)
canvas = tk.Canvas(right_frame, bg="grey", highlightthickness=0)
canvas.pack(side="left", fill="both", expand="true", padx=30)
scrollbar = tk.Scrollbar(right_frame, command=canvas.yview, width=20, troughcolor="#3E3B3C", bd=0,
activebackground="grey", bg="grey", relief="raised")
scrollbar.pack(side="left", fill="y")
scroll_frame = tk.Frame(canvas, bg="grey", pady=20)
scroll_win = canvas.create_window((0, 0), window=scroll_frame, anchor='nw')
canvas.configure(yscrollcommand=scrollbar.set)
canvas.bind('<Configure>', lambda event, canvas=canvas, win=scroll_win: on_configure(event, canvas, win))
tk.Label(scroll_frame, text="* Required Fields", bg="grey", fg="yellow", font="Arial 13 bold").pack()
input_fields = ("Name", "Mobile number","Email")
ip_len = len(input_fields)
required = [1, 1, 1, 1]
entries = []
for i, field in enumerate(input_fields):
print()
row = tk.Frame(scroll_frame, bg="grey")
row.pack(side="top", fill="x", pady=15)
```

```
label = tk.Text(row, width=20, height=1, bg="#3E3B3C", fg="#ffffff", font="Arial 13",
highlightthickness=0, bd=0)
label.insert("insert", field)
label.pack(side="left")
if(required[i] == 1):
label.tag_configure("star", foreground="yellow", font="Arial 13 bold")
label.insert("end", "  *", "star")
label.configure(state="disabled")
#if(i != ip_len):
ent = tk.Entry(row, font="Arial 13", selectbackground="#90ceff")
ent.pack(side="right", expand="true", fill="x", padx=10)
entries.append((field, ent))
tk.Button(scroll_frame, text="Register", command=lambda: register(entries, required, menu_var),
font="Arial 15 bold",
bg="black", fg="white", pady=10, padx=30, bd=0, highlightthickness=0,
activebackground="#3E3B3C",
activeforeground="white").pack(pady=25)
def startRecognition():
global img_read, img_label
if(img_label == None):
messagebox.showerror("Error", "No image selected. ")
return
crims_found_labels = []
for wid in right_frame.winfo_children():
wid.destroy()
frame = cv2.flip(img_read, 1, 0)
gray_frame = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
face_coords = detect_faces(gray_frame)
if (len(face_coords) == 0):
#messagebox.showinfo("Information", "Image doesn't contain any face or contains empty bike.")
now2=tk.Label(root,text="Normal Mode",font="times 22 bold",width=100)
now2.place(x=900,y=300,width=300,height=60)
```

```
wrong2=tk.Label(root,text="Normal",font="times 22 bold",width=100)
wrong2.place(x=800,y=400,width=500,height=60)
wrong2.after(2000, wrong2.destroy)
now2.after(2000,now2.destroy)
else:
(model, names) = train_model()
print('Training Successful. Detecting Faces')
(frame, recognized) = recognize_face(model, frame, gray_frame, face_coords, names)
img_size = left_frame.winfo_height() - 40
frame = cv2.flip(frame, 1, 0)
showImage(frame, img_size)
print("showImage")
if (len(recognized) == 0):
#messagebox.showwarning("Error", "UNKNOWN PERSON DETECTED.")
if current_hour >= 18 or current_hour < 6:
wrong=tk.Label(root,text="Unknown Person Detected",font="times 22 bold",width=100)
wrong.place(x=800,y=400,width=500,height=60)
now=tk.Label(root,text="Active Mode",font="times 22 bold",width=100)
now.place(x=900,y=300,width=300,height=60)
for i in range(1):
engine=pyttsx3.init()
engine.say("UNKNOWN PERSON DETECTED")
engine.runAndWait()
time.sleep(0.7)
print("fail")
messagebox.showwarning("Information", "UNKNOWN PERSON DETECTED.")
host = "smtp.gmail.com"
mmail = "egowthamhari2019@gmail.com"
hmail = "harilingapradeep2001@gmail.com"
receiver_name = "...."
sender_name= "Security Officer"
msg = MIMEMultipart()
```

```
subject = "'Hello,
'Someone try to access your bike',
For further Details
Contant us- 1098
Email-id: egowthamhari2019@gmail.com"'
text = "'Hello,
'Someone try to access your bike',
For further Details
Contant us- 1098
Email-id: egowthamhari2019@gmail.com"'
msg = MIMEText(text, 'plain')
msg['To'] = formataddr((receiver_name, hmail))
msg['From'] = formataddr((sender_name, mmail))
msg['Subject'] = 'Hello, my friend ' + receiver_name
server = smtplib.SMTP(host, 587)
server.ehlo()
server.starttls()
password = "kscbyldcbxtrvouz"
server.login(mmail, password)
server.sendmail(mmail, [hmail], msg.as_string())
server.quit()
print('send')
messagebox.showinfo('Email Sent'," succesfully Sent")
url='http://iotbegineer.com/api/sensors'
myobj={'sensor1':'Unknown Person Detected','sensor2':'Chennai','sensor3':'Active
Mode','sensor4':'Night','sms':'1'}
r=requests.post(url,json=myobj,headers={'username':'iotbegin370','Content-
Type':'application/json'})
print(r.text)
wrong.after(3000, wrong.destroy)
##          img_label.after(3000,img_label.destroy)
now.after(3000,now.destroy)
```

```
else:
wrong=tk.Label(root,text="Unknown Person Detected",font="times 22 bold",width=100)
wrong.place(x=800,y=400,width=500,height=60)
now=tk.Label(root,text="Normal Mode",font="times 22 bold",width=100)
now.place(x=900,y=300,width=300,height=60)
for i in range(1):
engine=pyttsx3.init()
engine.say("UNKNOWN PERSON DETECTED")
engine.runAndWait()
time.sleep(0.7)
print("fail")
messagebox.showwarning("Information", "UNKNOWN PERSON DETECTED.")
host = "smtp.gmail.com"
mmail = "egowthamhari2019@gmail.com"
hmail = "harilingapradeep2001@gmail.com"
receiver_name = "...."
sender_name= "Security Officer"
msg = MIMEMultipart()
subject = "'Hello,
'Someone try to access your bike',
For further Details
Contant us- 1098
Email-id: egowthamhari2019@gmail.com"'
text = "'Hello,
'Someone try to access your bike',
For further Details
Contant us- 1098
Email-id: egowthamhari2019@gmail.com"'
msg = MIMEText(text, 'plain')
msg['To'] = formataddr((receiver_name, hmail))
msg['From'] = formataddr((sender_name, mmail))
def getPage3():
```

```
root.destroy()
tk.Label(pages[0], text="TWO WHEELER THEFT PREDICTION", fg="black",
activeforeground="#A0CFEC",bg="#A0CFEC",font="Arial 15 bold", pady=30).pack()
tk.Label(pages[0], bg="#A0CFEC").pack(side='left')
btn_frame = tk.Frame(pages[0], pady=30)
#btn_frame.pack(side='left', padx=100,pady=10)
btn_frame.place(x=170,y=150)
tk.Button(btn_frame, text="REGISTRATION", command=getPage1, highlightthickness=0, bd=0)
tk.Button(btn_frame, text="SURVEILLANCE", command=getPage2, highlightthickness=0, bd=0)
tk.Button(btn_frame, text="EXIT", command=getPage3, highlightthickness=0, bd=0)
for btn in btn_frame.winfo_children():
btn.configure(font="Arial 12 bold", width=24, bg="#A0CFEC", fg="black", pady=1,
highlightthickness=0, bd=0,activebackground="#A0CFEC", activeforeground="black")
btn.pack(pady=30)
pages[0].lift()
root.mainloop()
```

A.2 SAMPLE SCREENS

FRONT PAGE

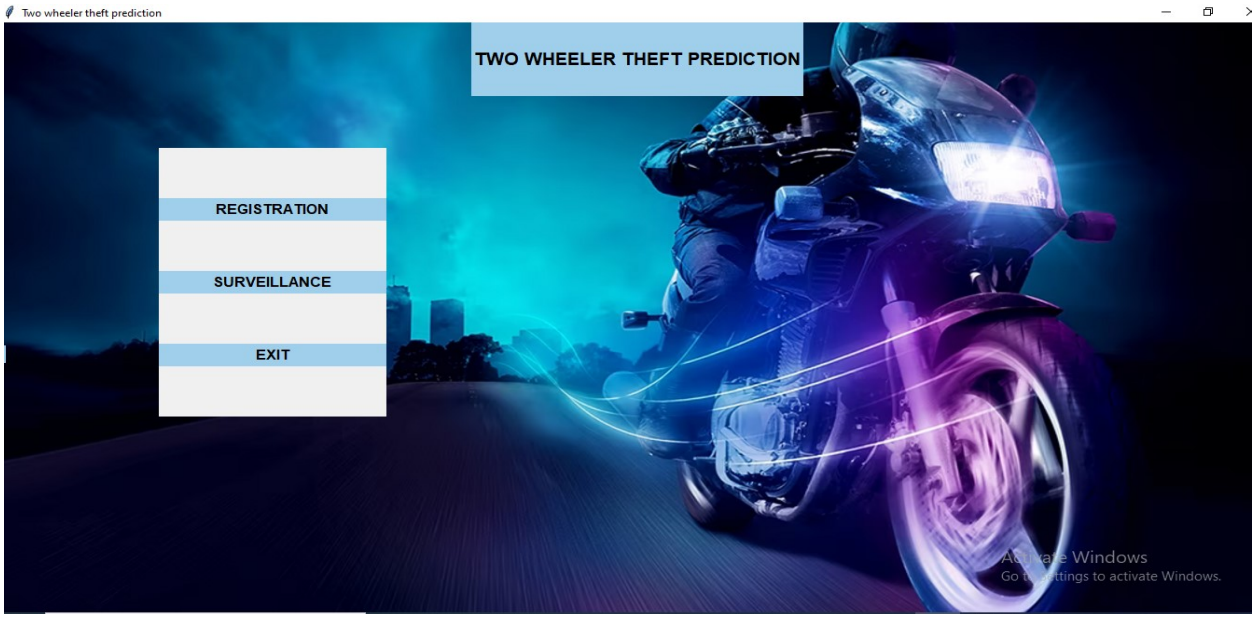

FIG 8.1 Front Page for Vehicle Theft Detection

REGISTRATION PAGE

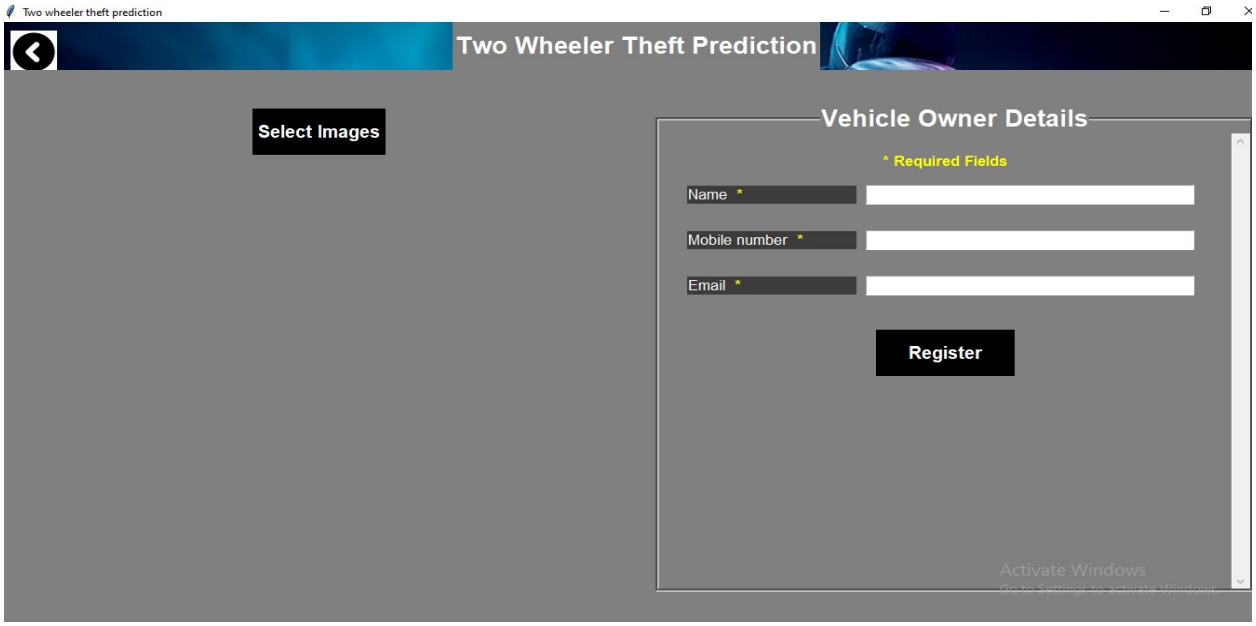

FIG 8.2 Registration Page for Vehicle Theft Detection

SURVEILLANCE PAGE

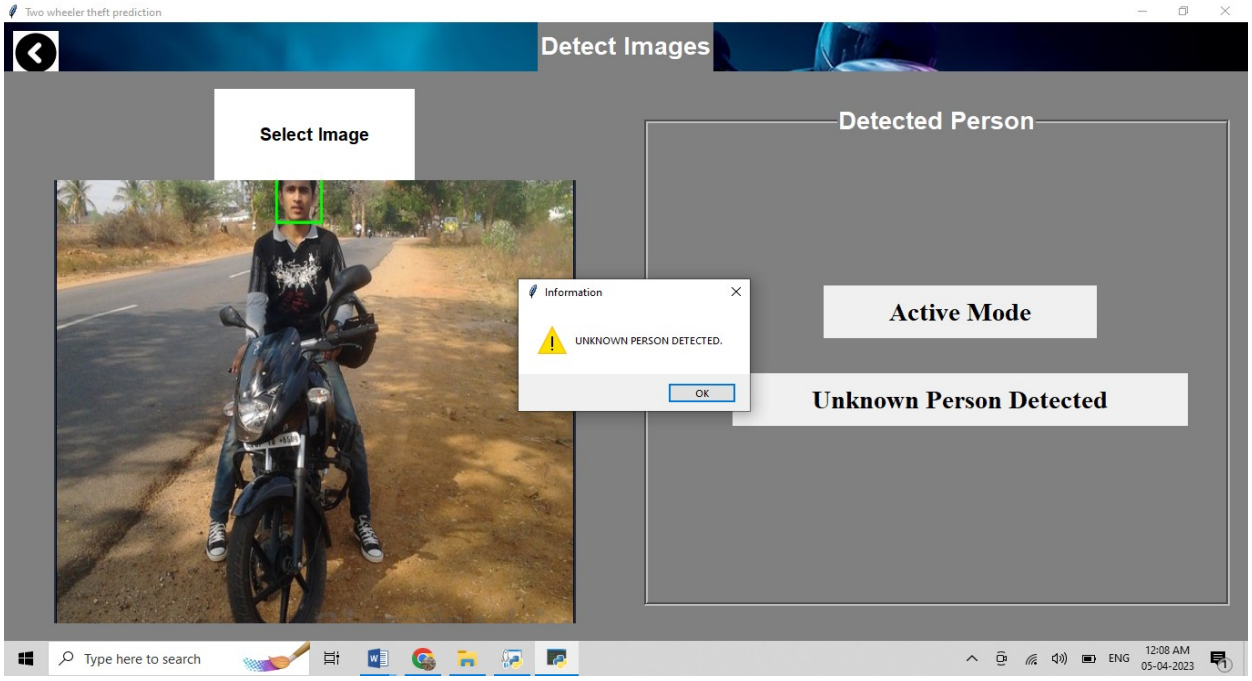

FIG 8.3 Before Registration of owner vehicle

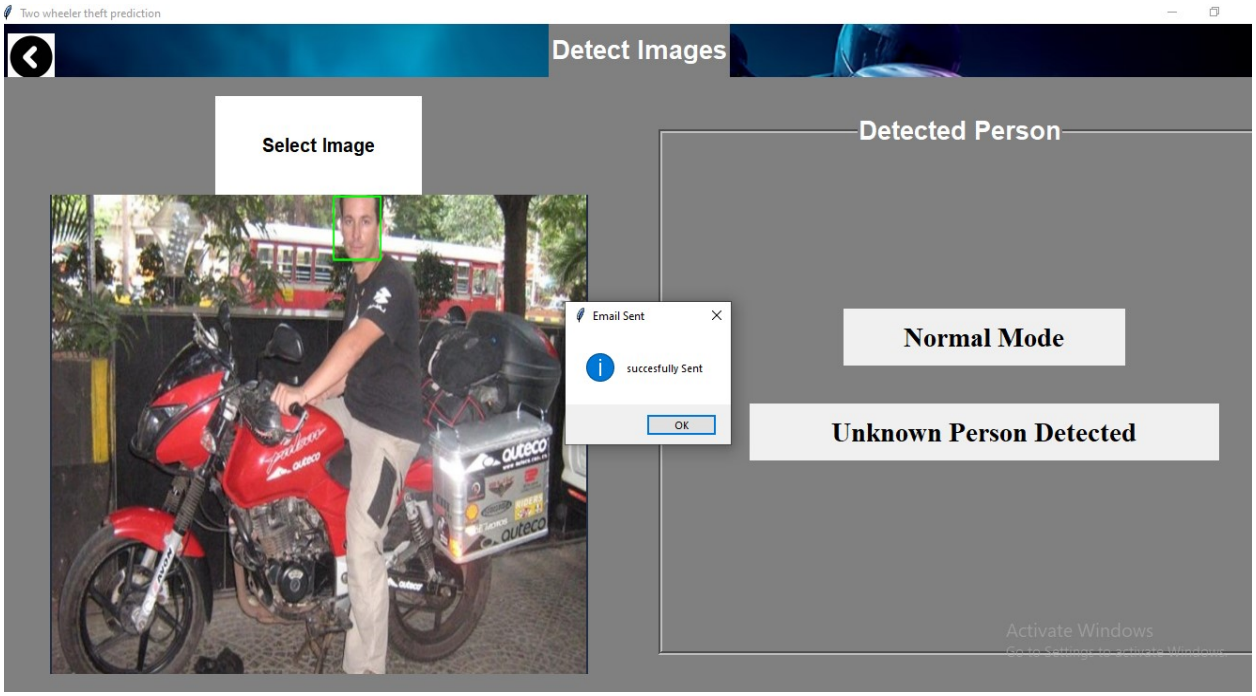

FIG 8.4 Surveillance image for unknown person

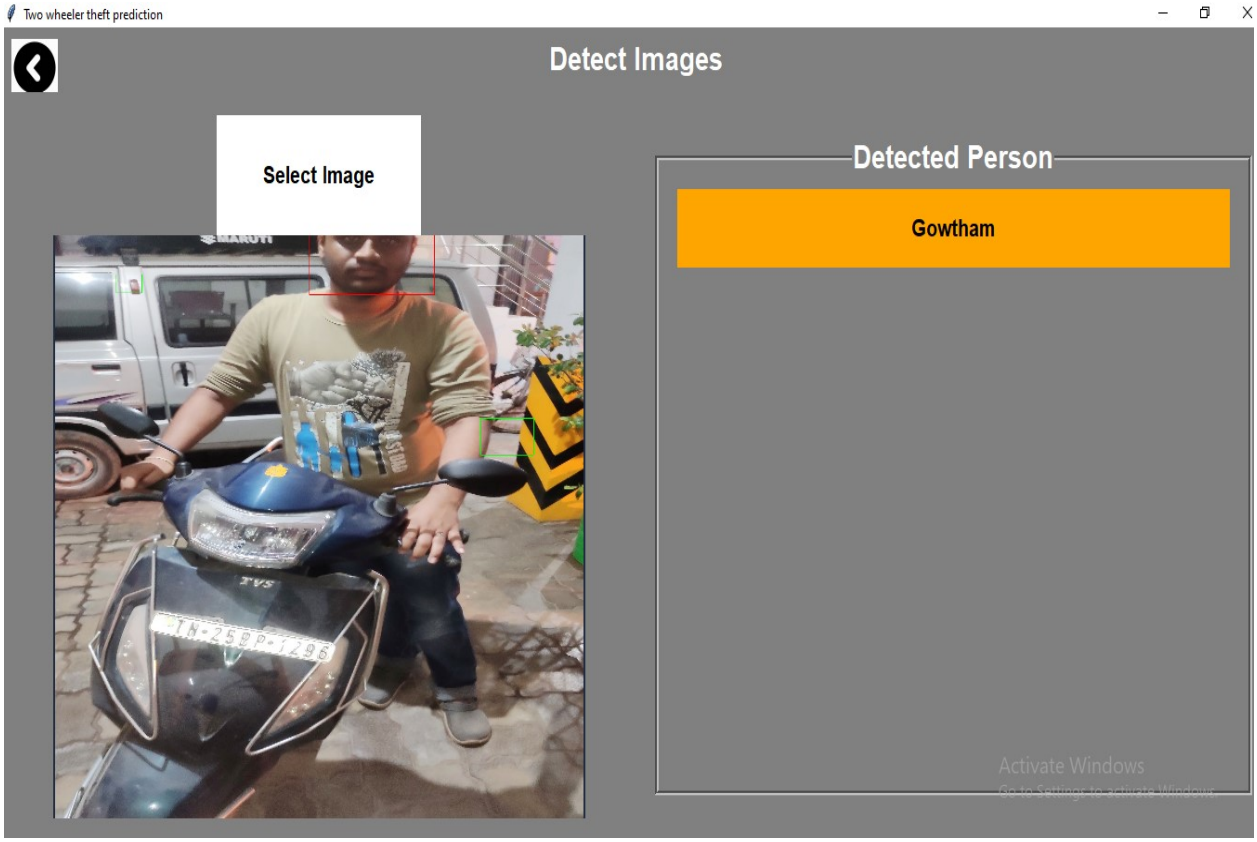

FIG 8.5 Surveillance image for known person

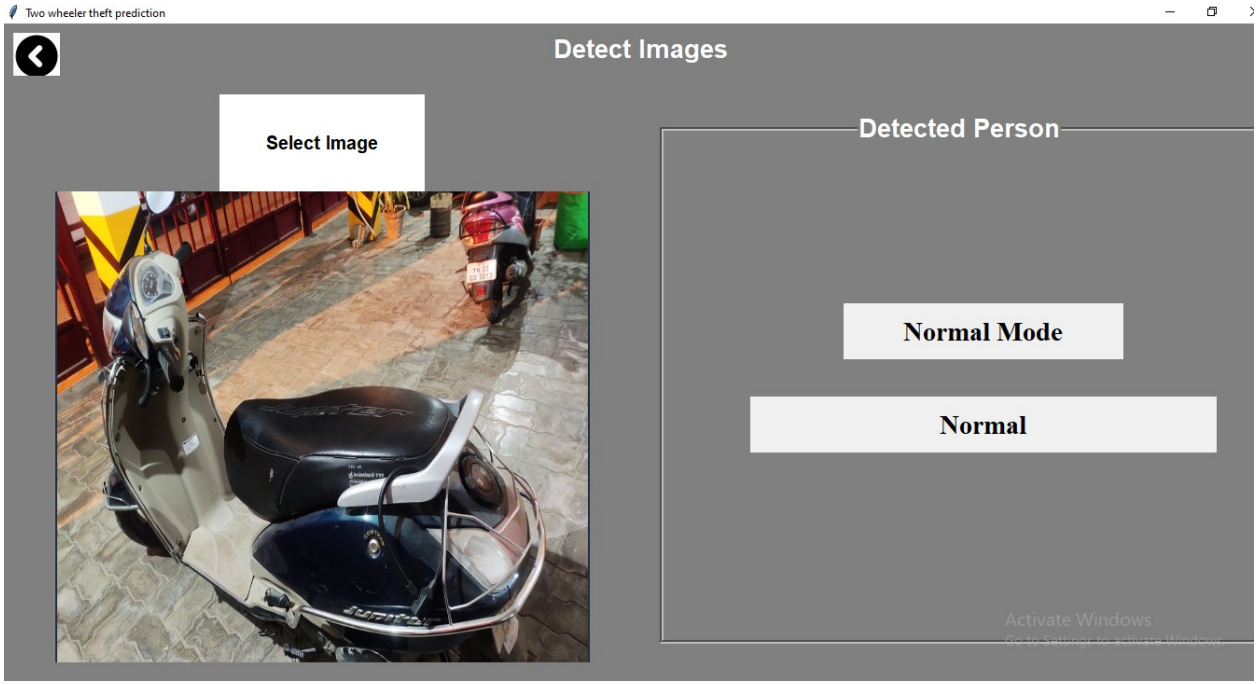

FIG 8.6 Surveillance image for without person

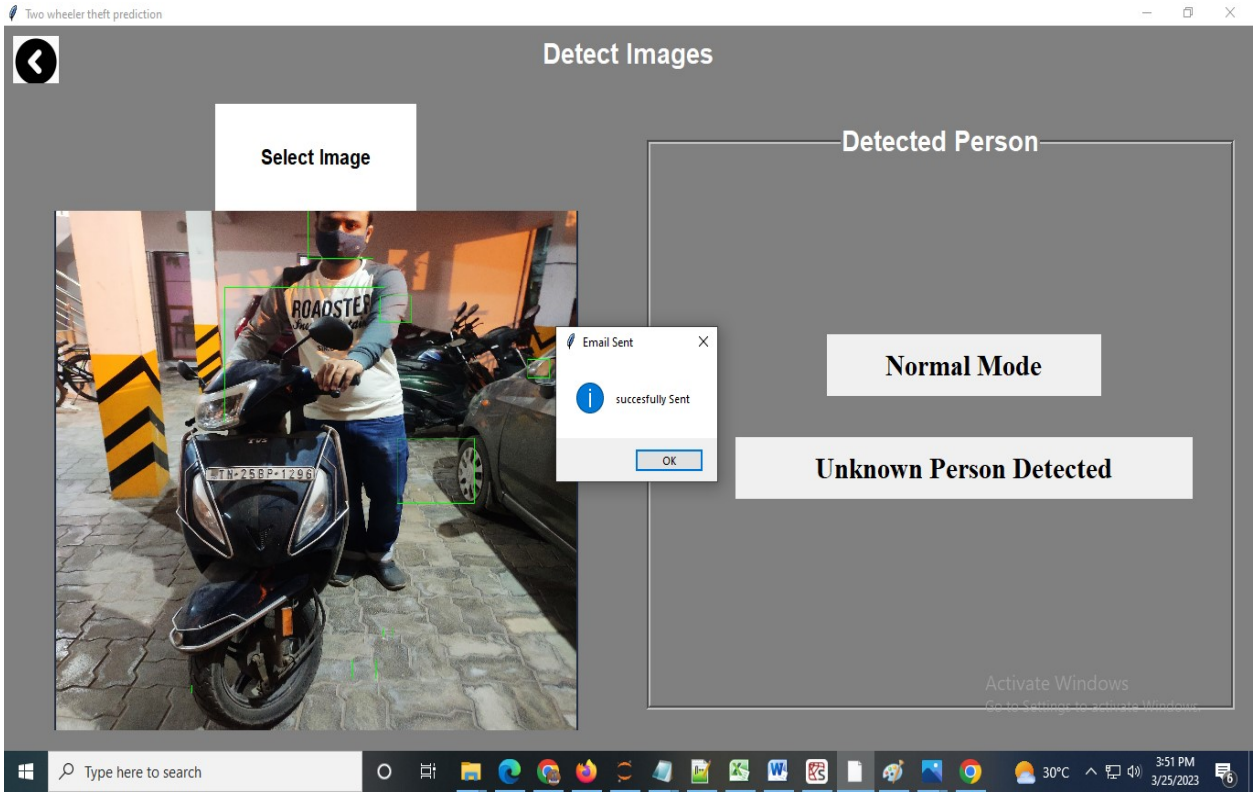

FIG 8.7 Surveillance image for person wearing mask

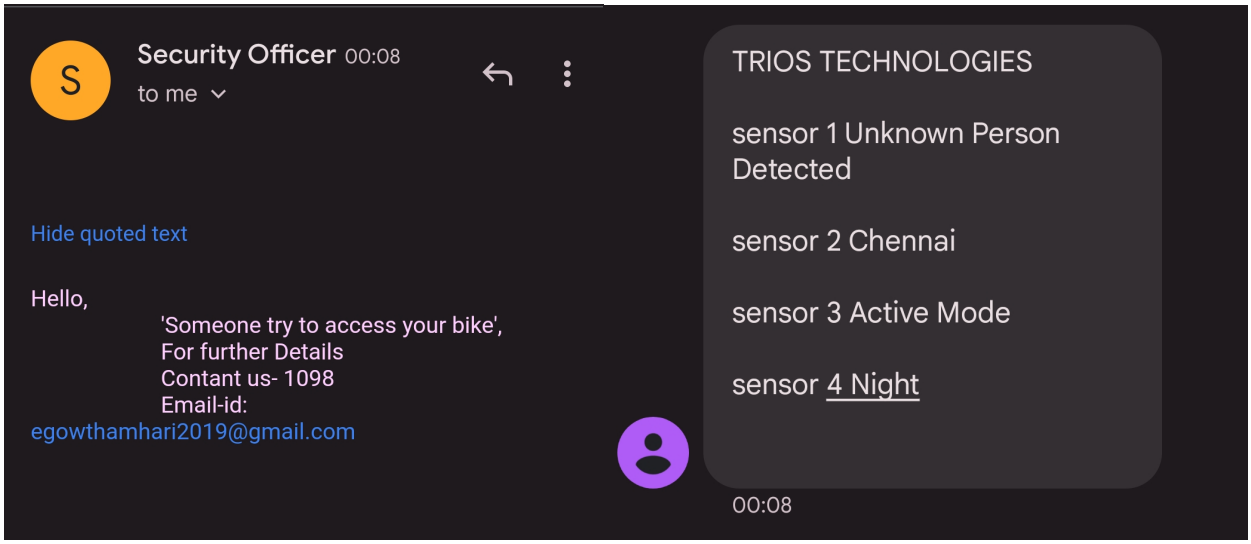

FIG 8.8 Message sent to owner vehicle

REFERENCES

[1] Z. Yan and H. Wen, ‘‘Performance analysis of electricity theft detection for the smart grid: An overview,’’ IEEE Trans. Instrum. Meas., vol. 71, pp. 1–28, 2022

[2] K. Fei, Q. Li, C. Zhu, M. Dong, and Y. Li, ‘‘Electricity frauds detection in low-voltage networks with contrastive predictive coding,’’ Int. J. Electr. Power Energy Syst., vol. 137, May 2022, Art. no. 107715.

[3] H. W. F. G. B. S. E. A. Arman Syah Putra, " “A Proposed surveillance model in an Intelligent Transportation System (ITS)”," 1st 2018 Indonesian Association for Pattern Recognition International Conference, INAPR, 2019.

[4] X. Yuan, P. He, Q. Zhu, and X. Li, ‘‘Adversarial examples: Attacks and defenses for deep learning,’’ IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 9, pp. 2805–2824, Sep. 2019.

[5] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, ‘‘A survey on security threats and defensive techniques of machine learning: A data driven view,’’ IEEE Access, vol. 6, pp. 12103–12117, 2018.

[6] N. Papernot, P. D. McDaniel, A. Sinha, and M. P. Wellman, ‘‘SoK: Security and privacy in machine learning,’’ in Proc. IEEE Eur. Symp. Secur. Privacy (EuroS&P), Apr. 2018, pp. 399–414.

[7] Y. Chai, Y. Zhou, W. Li, and Y. Jiang, ‘‘An explainable multi-modal hierarchical attention model for developing phishing threat intelligence,’’ IEEE Trans. Dependable Secure Comput., early access, Oct. 12, 2021, doi: 10.1109/TDSC.2021.3119323.

[8] R. Valecha, P. Mandaokar, and H. R. Rao, ‘‘Phishing email detection using persuasion cues,’’ IEEE Trans. Dependable Secure Comput., early access, Oct. 8, 2021

[9] I. U. Khan, N. Javeid, C. J. Taylor, K. A. A. Gamage, and X. Ma, ‘‘A stacked machine and deep learning-based approach for analysing electricity theft in smart grids,’’ IEEE Trans. Smart Grid, vol. 13, no. 2, pp. 1633–1644, Mar. 2022

[10] J. Moon, S. Jung, J. Rew, S. Rho, and E. Hwang, ‘‘Combination of shortterm load forecasting models based on a stacking ensemble approach,’’ Energy Buildings, vol. 216, Jun. 2020, Art. no. 109921