

# **TIEVS – FRAUD IMAGE IDENTIFICATION AND PREVENTION SYSTEM**

Sahabandu Kankanam Arachchige Kelmi Imashi Madhushani

IT18082548

BSc (Hons) in Information Technology Specializing in Information  
Technology

Department of Information Technology

Sri Lanka Institute of Information Technology  
Sri Lanka

October 2021

# **TIEVS – FRAUD IMAGE IDENTIFICATION AND PREVENTION SYSTEM**

Sahabandu Kankanam Arachchige Kelmi Imashi Madhushani

IT18082548

Dissertation Submitted in Partial Fulfillment of the Requirements for the BSc (Hons)  
in Information Technology Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology


Sri Lanka

October 2021

## DECLARATION

I declare that this is my own work, and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as article or books).

Name	Student ID	Signature
Madhushani S K A K I	IT18082548	

The above candidate has carried out research for the bachelor's degree Dissertation under my supervision.

Signature of the supervisor:

Date:

## **ACKNOWLEDGMENT**

I would like to highly appreciate the continued guidance and expressing my heartfelt gratitude to our supervisor, Ms. Manori Gamage (Lecturer -Faculty of Computing SLIIT), and co-supervisor, Ms. Suriyaa Kumari (LECTURE -Faculty of Computing SLIIT), for thier suggestions and assistance in composing this paper. And I would like to express my deepest appreciation to her for their unwavering support, encouragement, and enthusiasm during the research, as well as their insightful comments on how to make the 'TIEVS' system more efficient and productive. And also grateful to Sri Lanka Institute of Information Technology lecturers, research team members and family members for their ongoing guidance and encouragement and also Mr. Austin Reese, Mr. Baris Dincer, and the Kaggle team deserve special recognition for generating and making analysis-foundational datasets freely available to all scholars and tech enthusiasts.

## ABSTRACT

Because of the scarcity of physical magazines, online classified advertising has grown in popularity. This proposal is prominently focused on implementing analysis on conflicting fake image systems, targeting the items being on sale within our 'Tievs' online classified advertising platform, specifically on Cars. Nowadays, the public does not utilize tangible information sources such as newspapers, magazines, booklets, Leaflets, etc. With the development of advanced technology, these advertisements could reach a higher range of target audiences through online classified advertisement platforms, much more easily and securely. without needing to travel personally for long distances and exchanging money with other people, just to apply for or to receive the advertisement post(s), no need to waste our time and safety. Hence it has become the goal of our 'Tievs' online web application to provide such facilities, all the while enhancing them to outsmart other products with similar intentions. Most classified advertising products currently in use do not prioritize detecting and preventing fake advertisement images or characteristic details from being submitted, let alone inspect them internally and inform the respected users who are in the process of submitting the advertisement. Even more, many implemented online classified advertising systems, simply do not exhibit rich user interfaces for smooth functionality or promote quality user experience and user-friendliness, in consonance with the latest trending technologies. Supply the solution to this through the 'Tievs' online advertisement platform, In the event, a customer is in the process of filling out the necessary details and attaching relevant images to their advertisement on our 'Tievs' online classified advertisement platform, the system will internally examine for fake, inappropriate, misleading, mismatching images within. If detected, the system will swiftly prompt the customer before the advertisement is submitted, that an inaccurate image is being attached, so that the customer can take the corrective measures and eliminate the problematic scenario, regardless of it being done accidentally or deliberately. Using Convolutional Neural Network powered figure deception recognition system was introduced, which gained prodigious precision with righteous clarity in fraud detection and prevention. Hence, the proposed solution's objective of surpassing former classified advertising systems in delivering customer's necessities, using the most lucrative, timesaving, human-centric, and error-preventive approaches, was accomplished. It was affirmative by the positively responded questionnaire regulated among prospective users by the authors.

Key Words –Classified advertising, Image processing, fraud detection, machine learning, Convolutional Neural Network

## TABLE OF CONTENTS

DECLARATION .....	1
ABSTRACT .....	3
TABLE OF CONTENTS.....	4
LIST OF FIGURES.....	6
LIST OF TABLES .....	1
LIST OF ABBREVIATIONS .....	2
1 INTRODUCTION.....	3
1.1 Background .....	3
1.2 Literature Survey .....	7
1.3 Research Gap.....	14
1.4 Research Problem.....	16
1.5 Main Objectives.....	<b>Error! Bookmark not defined.</b>
1.6 Sub Objectives.....	20
2 METHODOLOGY .....	21
2.1 High-Level System Design.....	21
2.1.1 Data source, tools, and technologies .....	22
2.1.2 System Diagram.....	23
2.1.3 Software development life cycle .....	<b>Error! Bookmark not defined.</b>
2.1.4 Implementation Stage.....	<b>Error! Bookmark not defined.</b>
2.1.5 Project Requirements.....	<b>Error! Bookmark not defined.</b>
2.2 Commercialization Aspects of The Product.....	27

2.3 Implementation & Testing.....	27
3 RESULTS & DISCUSSION .....	<b>Error! Bookmark not defined.</b>
3.1 Results.....	27
3.2 Research Findings.....	27
3.3 Discussion.....	<b>Error! Bookmark not defined.</b>
4 CONCLUSION .....	<b>Error! Bookmark not defined.</b>
5 REFERENCE .....	39
6 GLOSSARY .....	40
7 APPENDICES .....	41

## LIST OF FIGURES

Figure 1.1 - Add hilarious Fake Car Ads.....	6
Figure 1.2 - Classification of image forgery detection.....	7
Figure 1.3 Testing image of SURF algorithm.....	7
Figure 1.4 - SVM classifier to distinguish.....	11
Figure 1.5- detection accuracy .....	12
Figure 1.6- The curves of the validation accuracy .....	13
Figure 1.7- The structure of the proposed common fake feature network .....	15
Figure 1.8- Fake Car images ad in Craigslist .....	<b>Error! Bookmark not defined.</b>
Figure 2.1- Fake Car images ad in Craigslist .....	20
Figure 2.2- Overall high-level system architecture .....	23
Figure 2.3- Software Development Life Cycle.....	26
Figure 2.4- Commercialization Aspects.....	28
Figure 2.5- Train the best parameters to the CNN mode.....	29
Figure 2.6- Shuffle the dataset .....	30
Figure 2.7- Implementation of vehicle and non-vehicle.....	31
Figure 2.8- Postman of non-fraudulent image detection .....	32
Figure 2.9- Postman of fraudulent image detection.....	33
Figure 2.10-UI Implementation.....	20



## **LIST OF TABLES**

Table 1.1 Comparison Research Study .....	14
---	----

## **LIST OF ABBREVIATIONS**

ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
UI	User Interface

# **1 INTRODUCTION**

## **1.1 Background**

Presently, the utilization of online platforms such as Craigslist, Olx, Quikr, Carewale, CarDheko, Facebook Marketplace, ikman.lk, etc., for classified advertisement interaction, has become increasingly popular compared to its predecessor periodicals, radio, and television communication channels. This inevitability, along with rapid technological growth and digitalization, has made it quite lucrative for both sellers and shoppers in their own endeavors whilst using the evolving marketplace of online classified ads [1]. Regardless of numerous implementations, a remarkably enhanced application that incorporates complex technologies to efficiently derive prominent functionalities that facilitate customer expectations and requirements has yet to arise. It is undoubtedly troublesome when interactions happen with an application that is not much accommodating to consumer needs appropriately.

A large number of individuals have been victims of picture fraud in our modern era. Many individuals utilize technology to alter photographs and present them as evidence to deceive the court. To put an end to this, all photos published on social media should be correctly classified as real or false. Social media is a fantastic tool for socializing, sharing, and spreading knowledge; yet, if caution is not taken, it has the potential to mislead individuals and even wreak chaos due to accidental false propaganda. While the majority of the altered photographs are plainly manipulated owing to pixelization and amateurish work, some of them do seem real. Manipulated image, especially in the political arena, may build or shatter a politician's confidence.

Nonetheless, there are certain circumstances in which consumers are not quite capable of completing a selling price value for the product that they wish to market and will likely continue to do significant study on how to recognize and define a suitable pricing value. When engaged in pricing clarification, a great deal of work must be put into the process, sacrificing important time once again, and as a result, consumers may feel somewhat sluggish to continue with the sale of the goods entirely. This withdrawal concept will, consequently, result in a decrease in meeting the customer's primary

requirements and necessities provided by the specific classified advertising program. As a result, the portal's reputation and competition with other comparable systems may suffer. Although the ultimate goal of our research project is to launch the Tiefs application after extensive design and development to cater to various types of classifieds in the future, as well as more upgrades relevant to each criterion, on behalf of the scope of this project, which is to be completed within 12 months, we have decided to confine our research implementations to solely 'Car' classifieds.

The global market for used or second-hand automobiles has been expanding, and it is possible that it has more than doubled in the last several years as a result of the Covid-19 pandemic. However, like with many previous unexpected pandemic achievements, the period of uncertainty and personal sorrow has resulted in a surge in the selling and purchase of used automobiles. With a scarcity of new autos from automakers able to access older vehicles by April and May, and customers more cautious to spend in major items, used vehicle sales skyrocketed. In the preceding 18 months, the latest two months, August, and September, saw the highest increase rate in used automobile sales volume.

Most people understand the potential of data technology and form successful alliances to figure together to use technology to form residents' lives easier, healthier, and more productive. Therefore, all people using these online platforms for daily life activities will make their work easier. There, sellers, or advertisers sell their second-hand cars using some smart platforms. There, they upload different car images and other non-vehicle images also upload to the system. In this era, Smart platform systems cannot be recognized images that are fake or original. but the average person should be able to identify whether they are fake or irrelevant. Or they may be fooled into buying the product. People cannot detect photo forgeries. Therefore, when uploading fake images with image size, price tag, logo tags to the site an advertiser accidentally or deliberately, damages the user's view of the ad as well as the advertising firm. Then, Customers will be unsatisfied with the online platforms even, also company reputation will be lost. Therefore, company income is also lost.

Images are more user-friendly to human eyes than content. For instance, most people focus on the content of the image while watching the ads regardless of the context since images can talk and tell the story [2]. However, in the digital era manipulating images is not easy to detect and prevent. The main problem happens when conflicting images have spread the web without any control. People tend to believe that each picture they see is 2 real without considering the likelihood that those images could be tempered, faked, or control. All mentioned issues arise the need for further research in this field.

Fraudulent image detection has been investigated for many years. generally, fake image detection investigates different characteristics of images and attempts to seek out traces to research. As mentioned above, most of the normal fraudulent detection techniques are often categorized into three classes, copy-move detection, splicing detection, and image retouching detection.

Most digital images are saved in JPEG format. The image is split into non-overlapping blocks of  $8 \times 8$  pixels in JPEG compression. The discrete cosine transform (DCT) is evaluated and quantized for each block using a typical quantization matrix. Because the block discrete cosine transform (BDCT) is the primary operation in JPEG compression, any tampering with the picture causes a disruption in the local statistics of the block DCT coefficients. Changes in the local statistics of block DCT coefficients can be detected and utilized to determine the existence of forgeries.

Current forensic procedures necessitate the use of an expert to assess the veracity of a picture. We created a system that uses machine learning to evaluate whether a image is fake or not, and we made it available to the general public. This study will be divided into three sections, the first of which will focus on the implementation details, the second on the experimental results, and the third on the conclusion.



*Figure 1.1 - Add hilarious Fake Car Ads*

## 1.2 Literature Survey

Due to inadequate existing research specifically covering the overall proposed solution, previous studies vaguely alluding to the main component of this research are being discussed. The scientific literature on existing categorized platform image identification and prevention has been sparse. As a result, this section will discuss various types of research that have been conducted in recent years regarding Fake image identification, some of which were phenomenal, revealing results about the accuracy and efficiency levels of ML algorithms and Image processing algorithms that have been used for prediction.

Yuanfang Guo et al. [18] presented two simple yet powerful histogram and encoding-based detection techniques for false colorized pictures, which outperform several state-of-the-art colorization systems. However, their findings show that the efficacy of their system decreases when the testing and training pictures are created using different colorization methods or datasets.

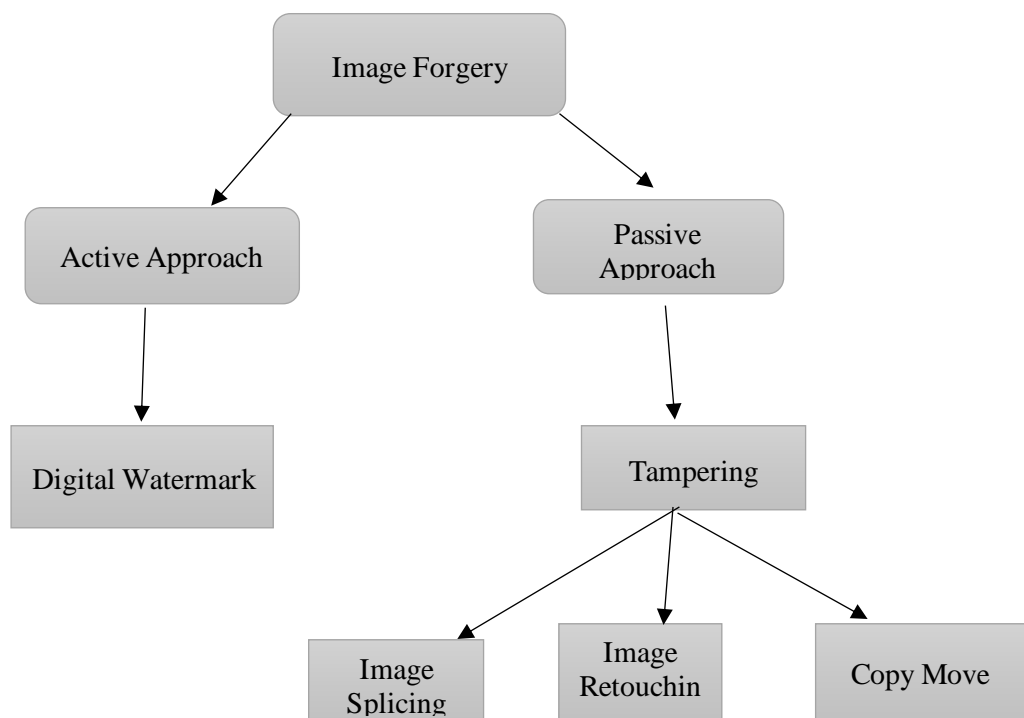


Figure 1.2 - Classification of image forgery detection

### **1.1.1. Copy-move detection**

Copy-move detection relies on identifying duplicated regions during a tampered image. Intuitively, these techniques tend to hunt an appropriate feature during a particular domain, such the detection are often performed via searching the foremost similar two units (such as patches). Different methods usually exploit different features.

Based on this technique, a basic and somewhat generic detection algorithm may have the following stages.

1. computing a dense NNF
2. field segmentation into areas defined by homogenous displacement vectors.
3. Pairs of candidates matching areas are chosen.
4. Suitable candidates are selected based on matching error and other criteria.

### **1.1.2. Splicing Detection**

In most regions using a variety of traces (features), which normally expose differences between tampered and untampered regions. Splicing detection is currently divided into four categories based on their mechanisms: compression-based methods, camera-based methods, physics-based methods, and geometry-based methods. cases, splicing detection detects manipulated regions that originate from various source images. Unlike copy-move detection, these methods detect tampered.



The above-mentioned processing approach, which was originally proposed in [13], may thus be described in the following phases.

1. calculating the high-pass residuals
2. truncation and quantization
3. Feature extraction based on selected neighbors' co-occurrence matrices.
4. On the training set, create an appropriate classifier.

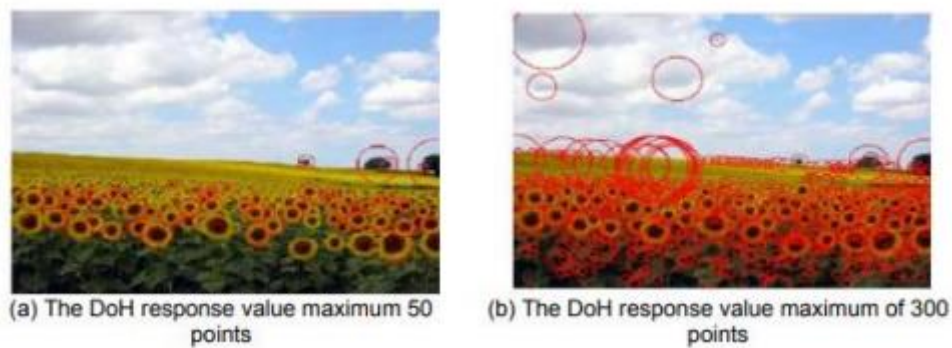
### **1.1.3. Image Retouching**

Detection In most cases, image retouching detection assumes that the original images have been restored or updated. for example, considers the similarities, distances, and number of equal pixels among different blocks to discern in painted pictures.

Similarly, [19] proposed a unique rapid image matching methodology that creates the k-d tree and employs improved BBF algorithms to replace the Linear algorithm and tests the performance of the modified method. Similarly, a few web-based scam detection programs, such as Advertising-Guard [20], aid inexperienced users in determining the validity of certain ads, therefore assisting law enforcement inspectors in averting online classified ad crimes.

[20] utilized Application Program Interfaces (APIs) and Image Recognition to evaluate the chance of a certain site being a scam by putting its URL into the tool. It did, however, have some limitations because it could only accept Craigslist advertising. Beneficial safeguards on fraudulent content/images have not been filled inside existing classified advertising systems, prompting this research to create a desirable technique to perhaps enhance application confidentiality entirely.

Feng Qi et al: In the study, it provides an efficient algorithm based on SURF. The research offers an effective SURF-based algorithm (Speeded Up Robust Features). To find and accurately describe digital images, the procedure uses the SURF algorithm, firstly, the SURF feature detector when extracting images and matching feature points in the image. The character description of the individual feature point vector will then be calculated using the DAISY algorithm rather than the SURF algorithm. The fake matching points are discarded with the RANSAC algorithm in the 4 process of matching functional points. Finally, it will calculate the space geometrical transformation parameters between two images according to the remainder of the match point and thus complete the matching procedure.[8]



*Figure 1.3 - Testing image of SURF algorithm*

Shilpa Dua et al: They demonstrated a novel integrated picture forgery detection method. The main concept is to take advantage of the variance in statistical characteristics of the entire image's AC coefficients by computing the standard deviation and count of non-zero DCT coefficients corresponding to each AC frequency component individually. For the test image and its cropped counterpart, the indicated characteristics are assessed. The retrieved feature vector is then utilized in conjunction with the SVM classifier to distinguish between changed and unmodified images.

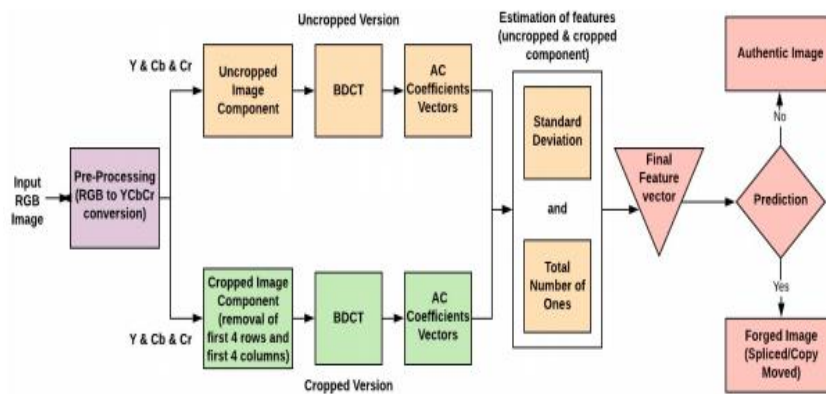


Figure 1.4 - SVM classifier to distinguish

in their experiment, Varying sets of test images with different sizes of changed sections, such as tiny, medium, and big, are assessed in their research for both types of forgeries.

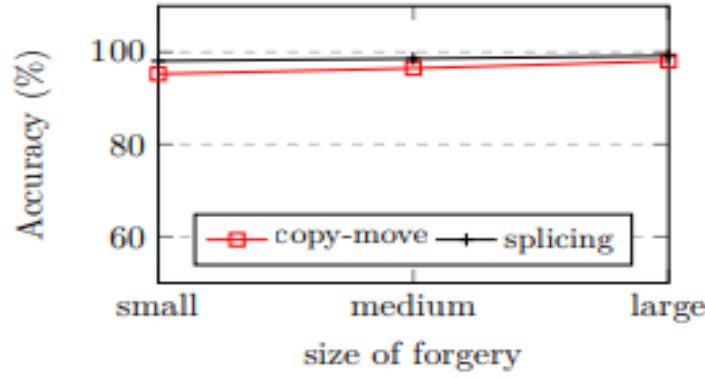


Figure 1.5- detection accuracy

Above Figure 4 illustrates the method's detection accuracy in the case of spliced and copy-moved pictures from their study image dataset. The figure clearly shows that the detection rates are consistent regardless of the size of the fabricated region.

Chih-Chung-Shu et al: To evaluate the efficacy of the suggested technique, they removed one of the chosen GANs from the training process and utilized it in the testing process instead, causing the training and test sets to be different. For example, when the PGGAN was removed from the proposed DeepFD's training phase, the misleading images created by the PGGAN and the matching actual images were utilized to assess the effectiveness of the trained fake face detector. Table 1 compares the objective performance of the proposed fake face detector, two baseline techniques, and methods offered in terms of accuracy and recall. As shown in Table 1, the suggested technique beat other state-of-the-art methods substantially; consequently, the CFFN may be utilized to capture the distinguishing aspects of the false pictures. Figure 4 depicts the validation accuracy curves throughout the training period. It proved the efficacy of the planned DeepFD. The suggested paired learning method effectively extracted CFFs from training pictures generated by various GANs. As a result, the proposed technique was shown to be more generalizable and effective than the other methods.

Method/Target	WGAN-GP		DCGAN		WGAN		LSGAN		PGGAN	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Method in [8]	0.322	0.373	0.334	0.349	0.371	0.391	0.350	0.396	0.345	0.378
Method in [9]	0.769	0.602	0.749	0.689	0.809	0.743	0.808	0.761	0.817	0.703
Method in [11]	0.792	0.684	0.820	0.811	0.864	0.881	0.848	0.869	0.868	0.853
Method in [10]	0.830	0.671	0.827	0.796	0.882	0.869	0.862	0.854	0.881	0.875
Method in [5]	0.832	0.690	0.871	0.847	0.885	0.920	0.866	0.898	0.922	0.909
Baseline-I	0.876	0.711	0.882	0.887	0.902	0.920	0.900	0.914	0.938	0.901
Baseline-II	0.901	0.728	0.822	0.838	0.864	0.881	0.920	0.919	0.917	0.887
The proposed	0.986	0.751	0.929	0.916	0.988	0.927	0.947	0.986	0.988	0.948

Table 1.1 - The objective performance comparison fake face detector

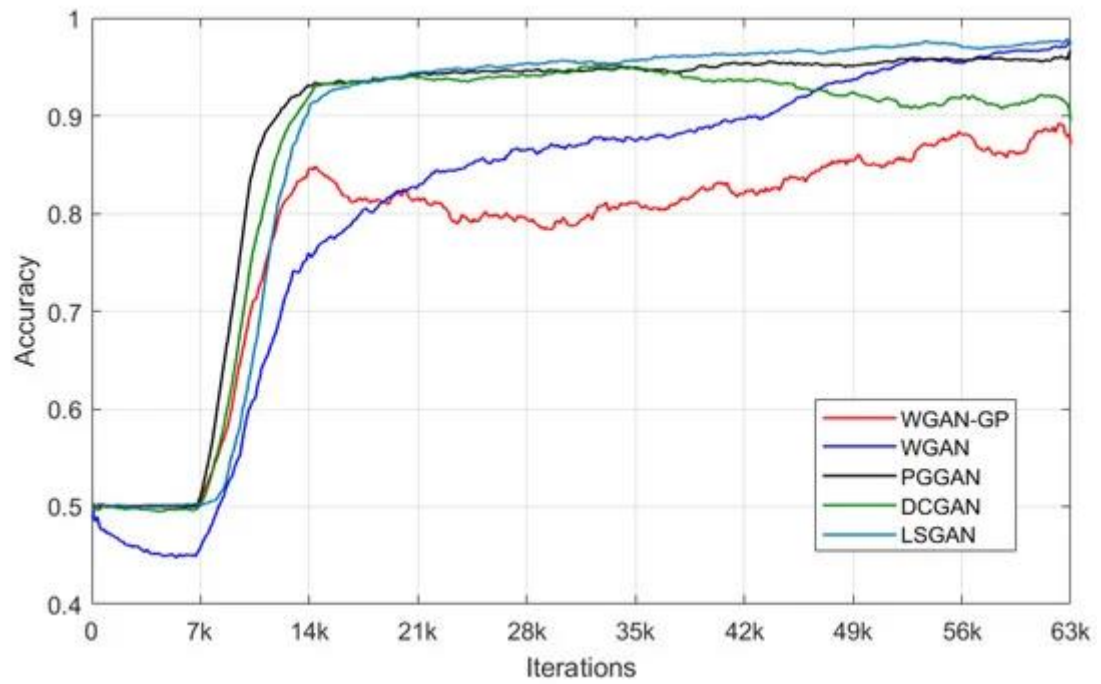


Figure 1.6- The curves of the validation accuracy

### 1.3 Research Gap

Even though It is perspicuous that some researchers have attempted to implement an outstanding model that is capable of detecting and prevention fake images with a decent accuracy level, still much more undiscovered innovative methodologies are yet to be realized.

The most common way of analyzing fake/irrelevant images is predicting whether fake or not for a section of text by doing a classification task. Moreover, former researchers were dependent on the use of some techniques for the detecting and preventing process. For example, Khaled A. N. Rashed et al [5] used the content-based image retrieval (CBIR) technique to detect fake images. And they used Web 2.0, Community and Collaborative Activities to analyze fake images. They created a system able to identify the fake images and image content, as well as the system, need to be able to manage user's interaction. They believe that content-based image retrieval (CBIR) techniques and harnessing collective intelligence is the appropriate solution f or this problem.

Chih-Chung Hsu et al: One of the most advanced CNN image identification car models is a dense block, a critical component of DenseNet. However, the monitored learning strategy is educated while the proposed CFF pair learning strategy refers to a semi

monitored learning strategy improved by them. A two-streamed network to enable CFF-learning feedback in a pair way is the proposed CFFN. The traditional CNNs, which are one-stroke, cannot obtain the pairing relevant information on the other hand, which means that the traditional CNNs cannot easily learn the common characteristics. In the CFFN suggested, any advanced CNN network, like ResNet Xception or DenseNet , may have been part of the backbone network. The performance of fake picture recognition will also be enhanced when the backbone network is trained to have the best feature representation capability. To that end, DenseNet is chosen by the proposed CFFN as a backbone network. [13]

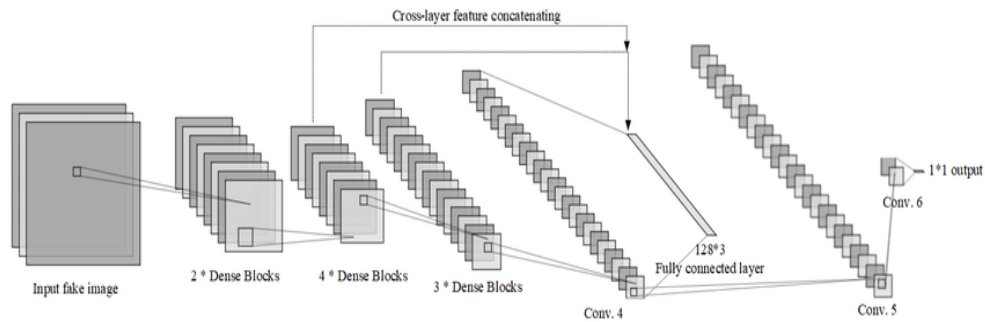


Figure 1.7- The structure of the proposed common fake feature network

Table 1.1 below depicts in tabular form, the summarization of the above explanation.

<b>Research</b>	<b>Addressing the cold start problem</b>	<b>Combination of different algorithm</b>	<b>Detection fake image</b>	<b>Prevention fake image</b>
<b>Research A</b>	×	✓	×	×
<b>Research B</b>	×	✓	×	✓
<b>Research C</b>	✓	×	✓	×
<b>Proposed Fraud Image Detection System</b>	✓	✓	✓	✓

*Table 1.2 Comparison of former research*



## **1.4 Research Problem**

User-friendly classified advertising platforms that provide the price predictions for selling products are obscure. In addition, from the applications' developers and owner's point of view, they have to implement strict program logic to identify or prevent fraudulent advertisements being posted (fake images/ irrelevant descriptions) accidentally or deliberately by users of the system, through verification and validation processes internally. This will require inspecting posts separately and meticulously, even after the ad being posted to the customers and if so, customers will view those fraud ads and lose confidence in using the platform for their needs, enabling loss of application reputation as well.

Most classified advertising products currently in use do not prioritize detecting and preventing fake advertisement images or characteristic details from being submitted, let alone inspect them internally and inform the respected users who are in the process of submitting the advertisement. Even more, many implemented online classified advertising systems, simply do not exhibit rich user interfaces for smooth functionality or promote quality user experience and user-friendliness, in consonance with the latest trending technologies.

From the perspective of the application's developers and owners, rigorous program logic must be implemented to identify or prevent fraudulent images from being posted (fake image/conflicting images) unintentionally or purposefully by system users, using internal verification and validation processes. This will need inspecting postings individually and thoroughly, even after the ad has been posted to consumers, and if this occurs, customers will see the fraudulent advertising and lose trust in utilizing the platform for their purposes, resulting in the loss of application reputation [2].



*Figure 1.8 - Fake Car images ad in Craigslist*

While it is obvious that many parties have already introduced online classified advertisement platforms in recent years, a highly improved application with a comprehensive machine-learning technology and Image processing algorithm has yet to be implemented to reinforce key functions in order to effectively meet customer requirements.

Furthermore, since Internet growth, these online ads have become the main part of the advertising market. And also, the internet advertising sites have a lot of advertising every day (Eg: eBay, Craigslist). Although advertisers had a high impact on the popularity of the online advertising sector. Therefore, it is a critical issue at present to proactively identify, identify or avoid fake images about the content of advertising. The widespread accessibility of the web attracts undesirable online scammers who pose as true sellers by posting fake ads with conflicting images a view to defrauding wishful buyers. Scammers can steal millions of dollars from unsuspecting users and threaten the reputation and utility of online ad services.

Recent approaches to image processing involve the instances of both fake pictures by user label or not. To enable an algorithm to learn how to categorize ads. I decided that use image processing to choose cases that are likely falsified images in order to overcome the implied problem. As a consequence, the proposed fraud image detection system attempts to overcome the above-mentioned frequent concerns of incomparable research studies, as well as create suitably unique modifications as answers to enhancing the user experience for the application's customers.

### **1.5 Main Objective**

- In this research, the last objective is to implement a Fraudulent image detection and prevention system for 'Tievs' classified advertising platform, Using a Deep learning model and Image processing algorithm.
- The main objective of this proposal is to implement the analysis of irrelevant / fake images posted on ads, and prompt a cautious message to the advertiser, using an image processing algorithm.
- This application supports the advertiser and the customer both. Most of the Advertising platforms not giving priority to Fraud image detection. And also, most of the available systems do not have a DL model to predict the fraud image. Therefore, while using the system by customers always getting wrong decisions depending on the fraud images or not. In this period, the Smart online advertising platform is the most suitable platform for buying and selling. Therefore, online fraudsters have also increased in sales ratio.

- A systematic and rich user-friendly smart solution to allow stakeholder advertising to be published efficiently and accurately without problems, as well as offering an optimized, interactive system search facility, to look and find items to buy seamlessly and productively in accordance with specific customers' requirements. (Will aim at one classified type of car, for example).
- ▪ The aim of this study is to implement a fake picture sensing system using an image processing algorithm to prevent an advertiser or persons with deceptive intentions from making the advertisement fake by themselves before submitting an advertisement for an item of a particular type (such as Cars) during the year.
- ▪ When the advertiser submits their advertisement into the system, is inspected by the images system it contains. If the advertiser added the fake images, the system would prompt the cautious message to the advertiser asking it to identify and remove this fake image.
- Ostensibly, the significance of an accurate, reliable, and efficient price prediction system for reselling cars is portrayed, thus the main objective of this proposed research study.

### **1.6 Sub Objectives**

- successful utilize of JPEG features algorithm, for image forgery detection.
- Proactively identifying and preventing fraudulent/ conflicting images, from being posted in the application.
- ▪ Identifying if the advertiser input ill-matched/ conflicting images into the application and prompts them cautious messages if they repeatedly attempt to input such irrelevant images.

## 2. METHODOLOGY

### 2.1 High-Level System Design

This industry will highlight how development is managed and carried out. A short conversation on the feasibility study is also included. As this is the implementation of a research project Agile development strategy is used to execute the system module by module.

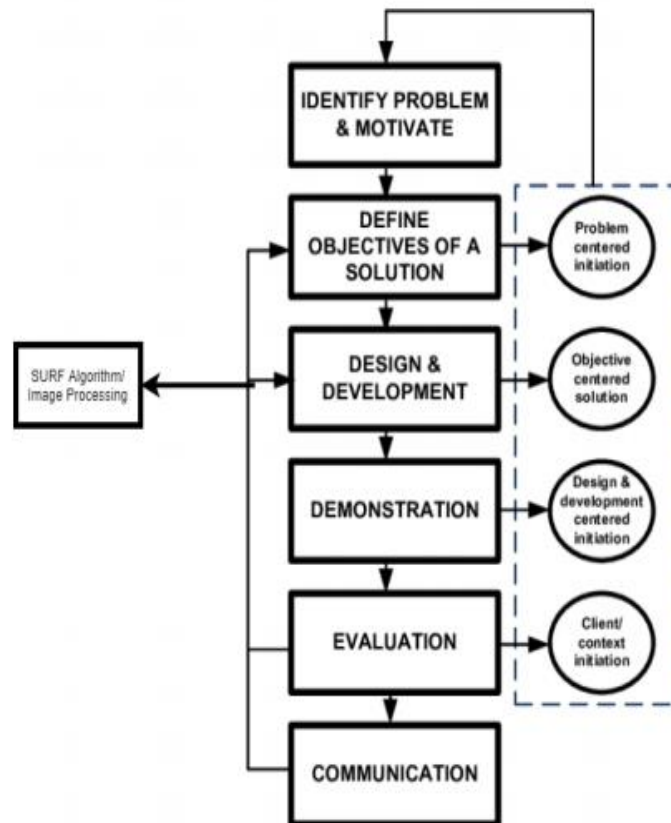


Figure 2.1 - Research Methodology diagram

### **2.1.1 Data source, tools, and technologies**

Two openly available Kaggle datasets were utilized for obligatory ML model rudiments. The ‘Used Cars Dataset’ sustaining 441,396 used car records covering about 40 brands having 25 unique attributes was scraped 2 months before this research inception from an American classified advertisement portal, namely ‘Craigslist’ since Tievs primitively targeted the US market. The ‘Vehicle Detection Image Set’ assisted the vehicle image recognition and fraud prevention strategy. Programming languages and frameworks as Python. Use Django and Django framework prominently,

- Jupyter Notebook
- Anaconda Navigator
- Visual Code Studio
- Keras and Tenserflor
- Numpy/Panda/Sklearn/Scikit Learn/ Matotlib/CV2/skimage.io
- NodeJS
- Pycharm
- Angular Framework

## 2.1.2 System Diagram

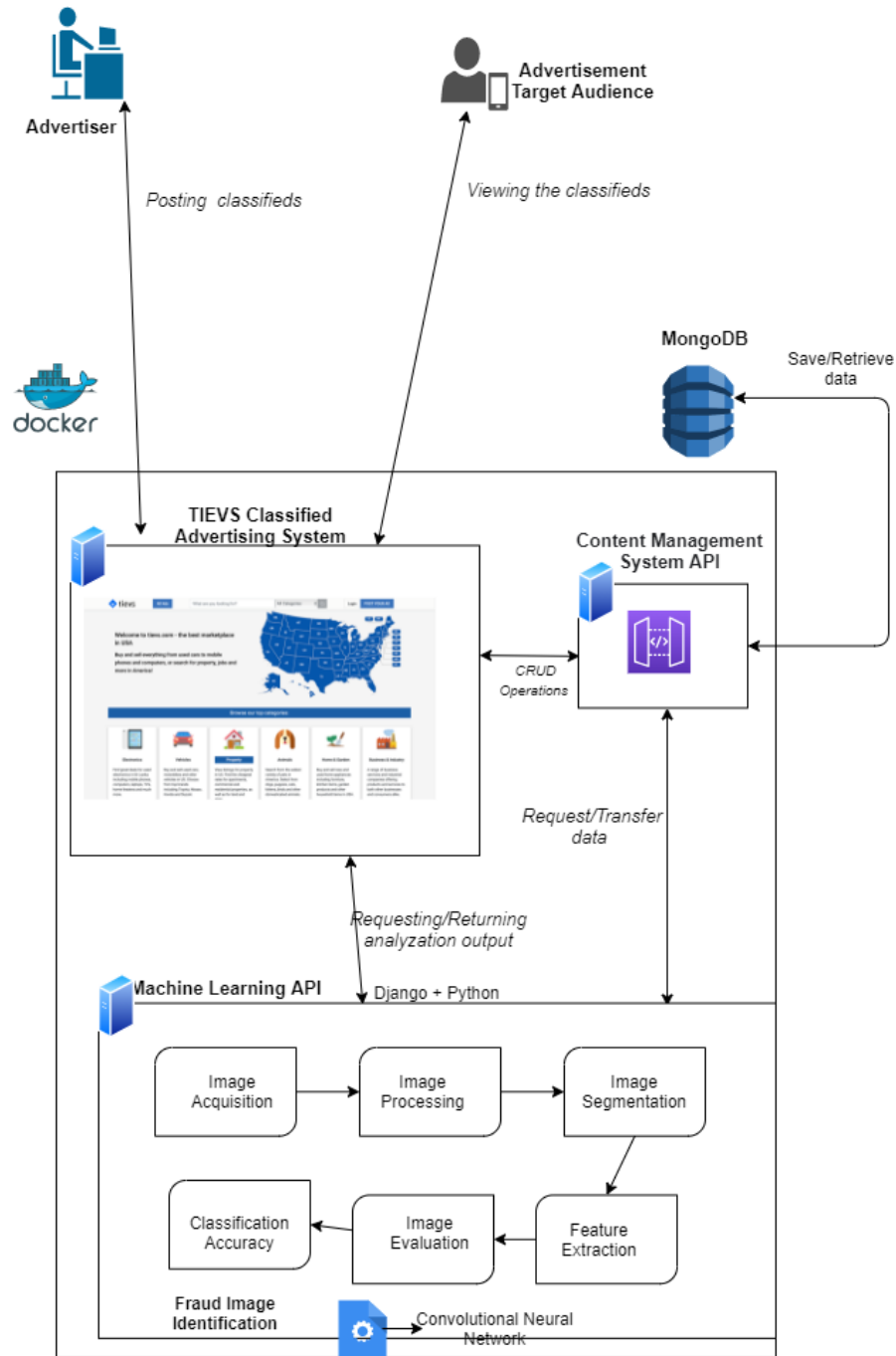


Figure 2.2 - Overall high-level system architecture

The above shown Fig. 2.2 depicts the Overall high-level system architecture of the newest component architecture built for Fraud image detection and prevention in 'Tievs' car classified advertising. The acquired dataset was first imported from the Django database into the Jupyter Notebook environment, where it was subjected to exploratory data analysis alongside various attribute selection techniques, feature engineering techniques, and other data preprocessing methods in order to understand and cleanse the dataset before transferring it into the models. The data was then divided into two primary subsets, X and Y, based on the independent factors and the dependent variable (vehicle). Following that, it was divided into several sets, such as x train, x test, y train, y test, using a split ratio for training and testing. A comparison study was carried out on Keras models in order to assess their effectiveness in binary prediction while using a complicated dataset. Those models were continually trained and tested using the same train and test datasets until satisfactory results were obtained.

The images will be extracted from the advertising form in the application interface and entered into the Django API, which contains the model, and the detection result will be generated after a thorough analysis of the features. Following the correctness of the model, it will be integrated with the Angular application to identify Fraud images of vehicle advertising that interact with the Frontend. The Advertiser will display the identification as unrelated images within the advertising form, near the images input area. Python is the main programming language to implement the system.



### 2.1.3 Software Development Life Cycle

An agile methodology is considered for the life cycle development of software. Furthermore, the Scrum methodology is primarily covered within the enormous Agile framework. Scrum is the lightweight framework that enables people, teams, and institutions to gain value for complex challenges by means of agile strategies. Product backlog creation, sprint planning and backlog creation, daily scrum conference, sprint retrospect, and sprint review are the incessant but iterative essential steps taken in a scrum strategy. The same methods will be pursued successfully for this proposed system. When you resume, it's safe to say. the same methods are successfully followed for this proposed system. It is safe to say that many unanticipated events will occur while research is carried out, particularly because some of the desired areas should be first analyzed. The existing and developed strategy will not be pursued the same way in this situation, because changes will have to be made when dealing with the impediments to the study process. In this way, agile methodology is important to make the implementation process easier and more consistent. Agile can provide research with increased control, improved efficiency, higher quality, satisfied customers, and greater return on investment according to Figure 2.3 below.



Figure 2.3 - Software Development Life Cycle (SDLC)

#### **2.1.4 Project Requirements**

##### ***Functional requirement***

- Successful customer navigation to the application
- The proper visual interface of advertising process elements should be implemented for a convenient representation of the different features of a car.
- Integration between the Deep learning API and the fake image analysis system.
- Proactively identify and prompt a warning message before submitting the advertisement

##### ***Non-Functional Requirements***

- Accuracy
- Scalability
- Adaptability
- Efficiency
- Ease of use
- Maintainability
- Reliability
- Less response Time
- Accessibility

## 2.2 Commercialization Aspects of The Implementation & Testing

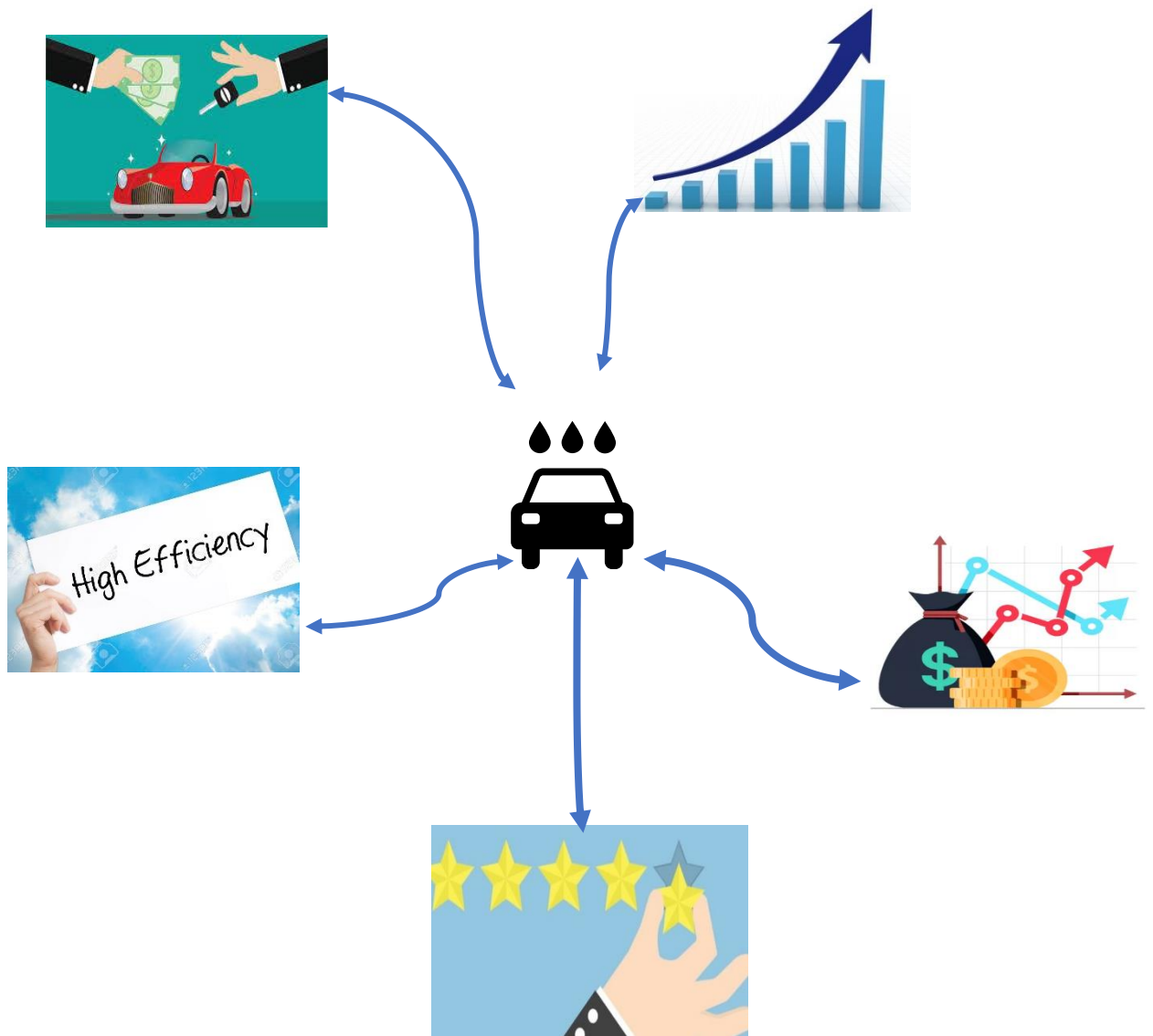
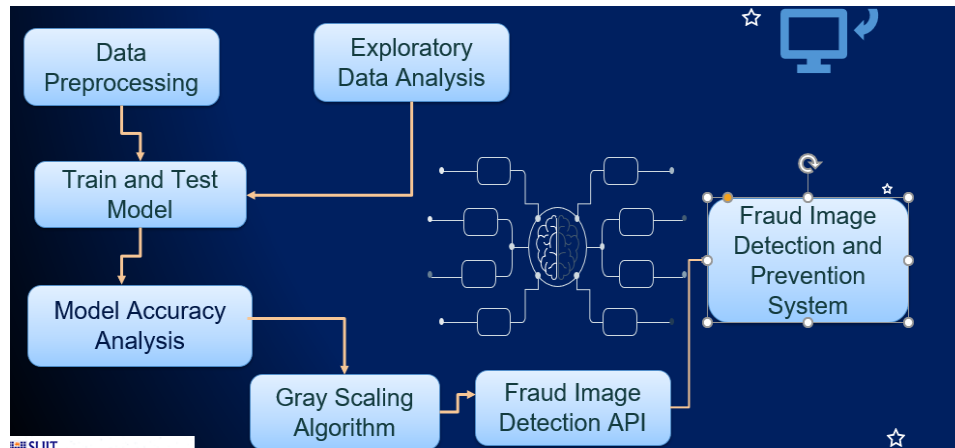


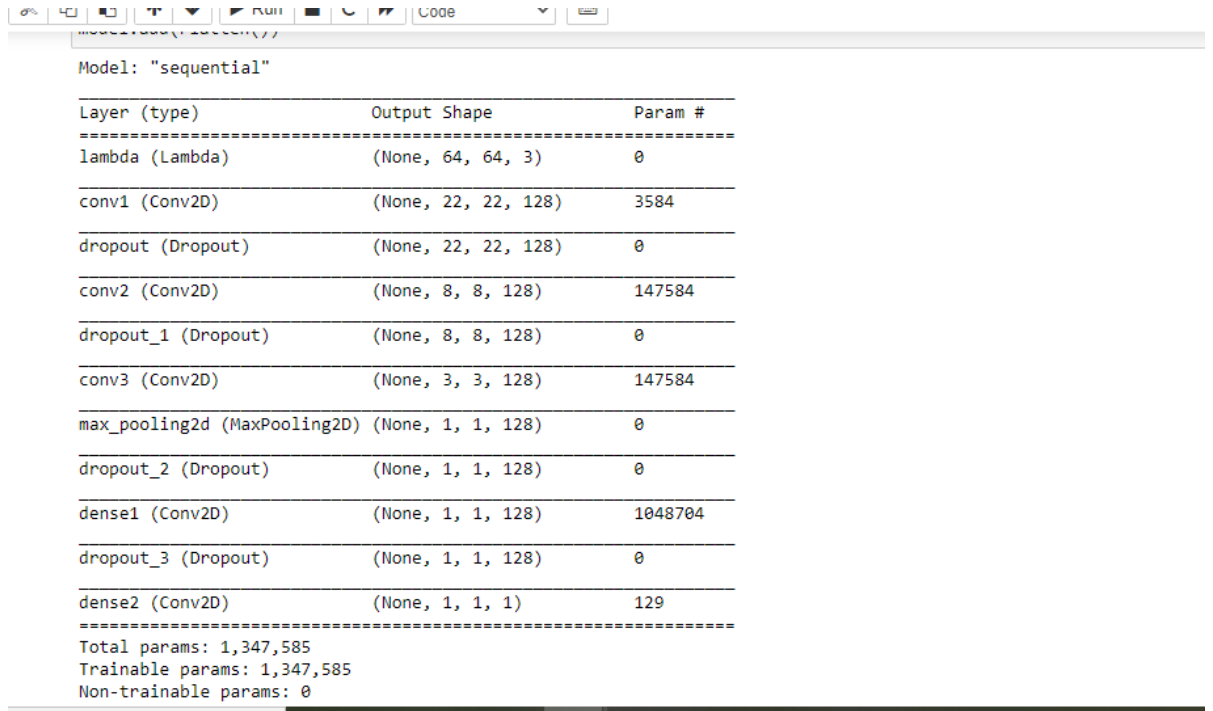
Figure 2.4- Commercialization Aspects

## 2.3 Implementation and Testing



### 2.3.1 Implementation

Authors stipulated the DL algorithm, **Convolution Neural Network (CNN)**, and in its convolutional layer, import the 10 layers, initial image feature learning was done using the vehicle/non-vehicle image set to extract characteristics such as gradient orientation, colors, edges, etc., assign weights and compress them dimensionally as convolved features compared to the original image. GREY scaling the data to execute image preparation procedures such as image rotation, resizing, and part extractions were done using Keras API. Next, the pooling layer reduced spatial size further while extracting superior traits and optimizing processing power. These 2 layers executed continuously and fastidiously trained the CNN model in discerning image qualities.



```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
lambda (Lambda)	(None, 64, 64, 3)	0
conv1 (Conv2D)	(None, 22, 22, 128)	3584
dropout (Dropout)	(None, 22, 22, 128)	0
conv2 (Conv2D)	(None, 8, 8, 128)	147584
dropout_1 (Dropout)	(None, 8, 8, 128)	0
conv3 (Conv2D)	(None, 3, 3, 128)	147584
max_pooling2d (MaxPooling2D)	(None, 1, 1, 128)	0
dropout_2 (Dropout)	(None, 1, 1, 128)	0
dense1 (Conv2D)	(None, 1, 1, 128)	1048704
dropout_3 (Dropout)	(None, 1, 1, 128)	0
dense2 (Conv2D)	(None, 1, 1, 1)	129

```

Total params: 1,347,585
Trainable params: 1,347,585
Non-trainable params: 0

```

*Figure 2.5 - Train the best parameters to the CNN mode*

The above Fig. 2.39 demonstrates the best parameters selected for the CNN model. The final output was flattened and column vectorized before classification through a fully connected feed-forward neural network layer. With backpropagation applied iterative training, the model could discriminate high/low-level image features preparatory to effectuating **Softmax classification**. After the model was thoroughly tested on the accuracy, it was capable of identifying ad permissible car images by the manufacturer as well as prohibited images. The specific ML API was created using the Django framework and integrated with the trained model before being implied into the front-end application to detect inappropriate images and display warning signs necessarily. CNN differentiates from the traditional model by using only a few parameters to reduce overfitting probability and considering neighbor context (images in this scenario) information to note similarities. Below Fig. 2.40 shows the importing of the vehicle/non-vehicle dataset.

```
In [7]: from sklearn.utils import shuffle
```

```
# shuffle dataset
dataset = shuffle(dataset)
dataset.head()
```

Out[7]:

	path	is_vehicle	image
2028	dataset/vehicles/2827.png	car	[[[19, 18, 13], [18, 17, 14], [16, 20, 16], [1...
10510	dataset/non-vehicles/extra2856.png	non_car	[[[69, 72, 85], [70, 73, 84], [65, 69, 79], [6...
12933	dataset/non-vehicles/extra5311.png	non_car	[[[176, 179, 184], [174, 177, 182], [172, 175,...
8007	dataset/vehicles/middle (368).png	car	[[[86, 94, 70], [83, 91, 67], [83, 92, 68], [8...
506	dataset/vehicles/1455.png	car	[[[19, 17, 13], [15, 15, 14], [14, 14, 12], [1...

Figure 2.6- Shuffle the dataset

In order to aim to predict the potential of fraudulent image submission prior to ad submission, an initial dataset record was categorized as related and non-related, with the latter being prepared. After successfully installing all of the packages (including libraries and dependencies) necessary for the operation, the exploratory data analysis (EDA) approach was largely used to get a communicative impression in the form of charts and graphs. Below shown Fig. 2.41 displays how the dataset is being divided into vehicle and non-vehicle sets.

```
In [4]: import pandas as pd
```

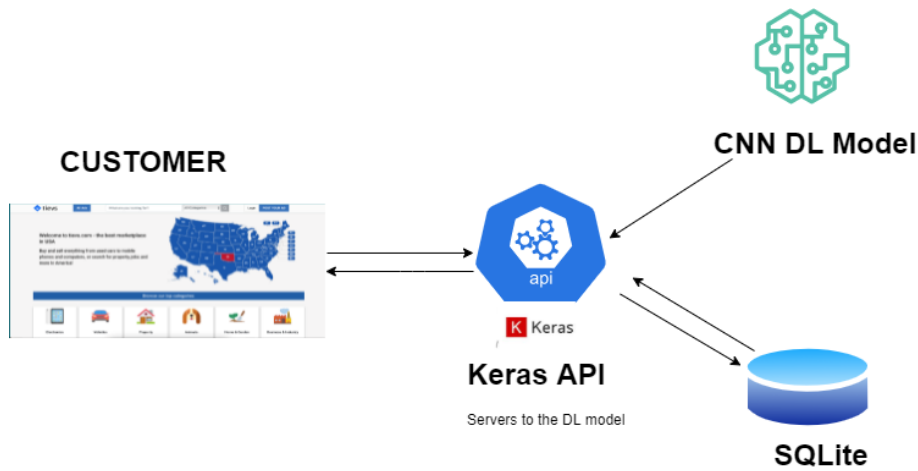
```
# add card to df
car_df = pd.DataFrame(cars, columns=['path'])
# add label 'car'
car_df['is_vehicle'] = 'car'
non_car_df = pd.DataFrame(non_cars, columns=['path'])
non_car_df['is_vehicle'] = 'non_car'

dataset = pd.concat([car_df, non_car_df], ignore_index=True, sort=False)

dataset.shape
```

Out[4]: (17760, 2)

Figure 2.7 - Implementation of vehicle and non-vehicle



### 2.3.2 Testing

During the implementation phase, all parts, from data engineering to front-end application deployment, underwent continuously and repeated testing. Tests were carried out concurrently with development activity and not simply at the end of the system's life cycle. Some of the functional testing types used throughout the construction of the fraud image detection system and Tiefs application as a whole included unit testing, smoke testing, component testing, integration testing, API testing, UI testing, black-box testing, and production testing. API testing of the CNN model integrated Django API was performed after the cycle to clearly and correctly evaluate the accuracy of image prediction. The purpose of this black-box testing process is to confirm the research components' compliance to their initial functional specifications and requirements.

Afterward, the confirmation of the model's function worked as expected, as well as determine the correctness of the algorithms and efficiency of the model and accuracy of the model Softmax classification and GRAY scaling were examined. The postman verified results are Showed. The Fig2.3.5 below shows the postman results in "non-fraudulent image detection" endpoint real-world input from the user after integrating and deployment for defining the purpose on the customer.

API POST endpoint: - <https://catsy.org/check-vehicle/>

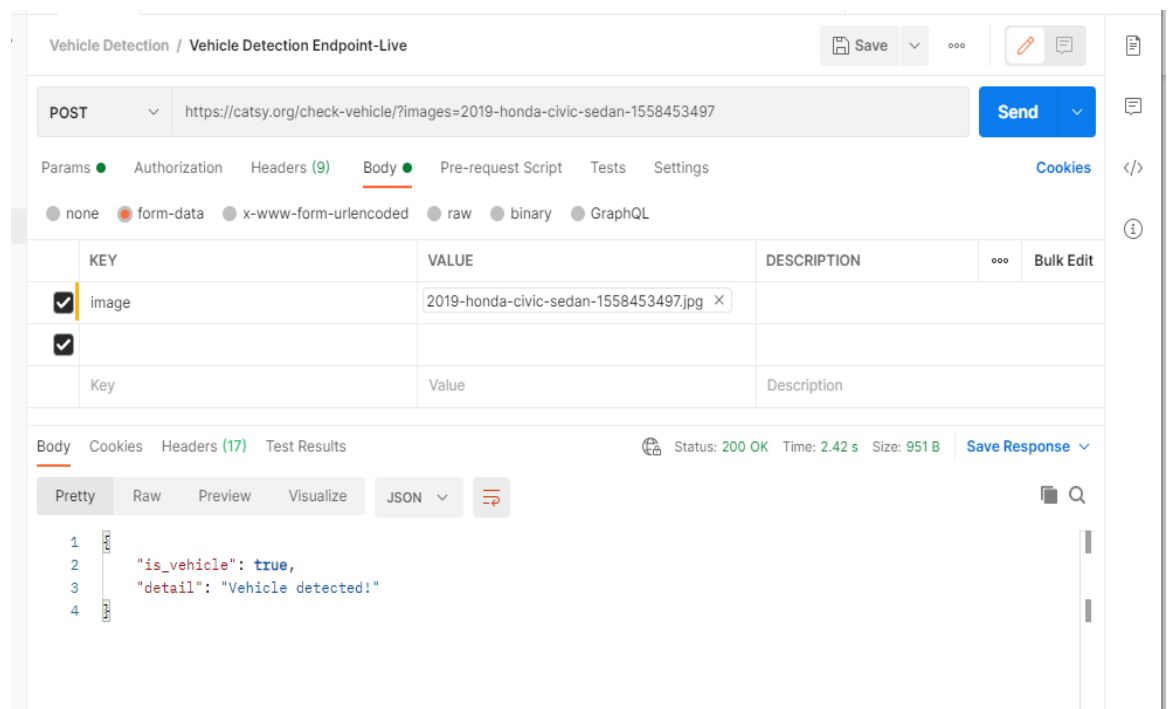


Figure 2.8 - Postman of non-fraudulent image detection

The below Fig 2.9 shows that successfully detected the fraudulent images in the system.



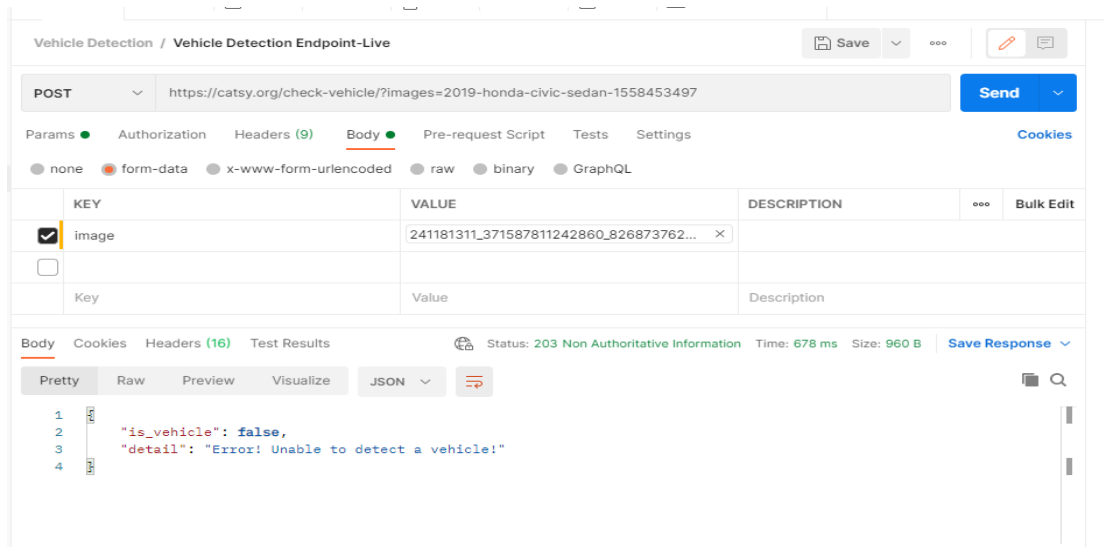


Figure 2.9 - Postman of fraudulent image detection

### 3. RESULTS AND DISCUSSION

#### 3.1 Results

The image recognition system was successfully able to detect 'fake' and 'not fake' images with minimal processing power and restrain inapposite images from being submitted into the Tievs application through deterrents. As given in Table 3.2, the CNN model accuracy rate is flattering, being above 90% with an adequate loss rate as well.

Table 3.1 Convolution neural network Statistics

Image Classification	Accuracy Rate	Loss Rate
Convolutional Neural Network	51.16%	0.8363

Accuracy level has obtained as shown in the Fig. 3.16, after training the model using preprocessing data.

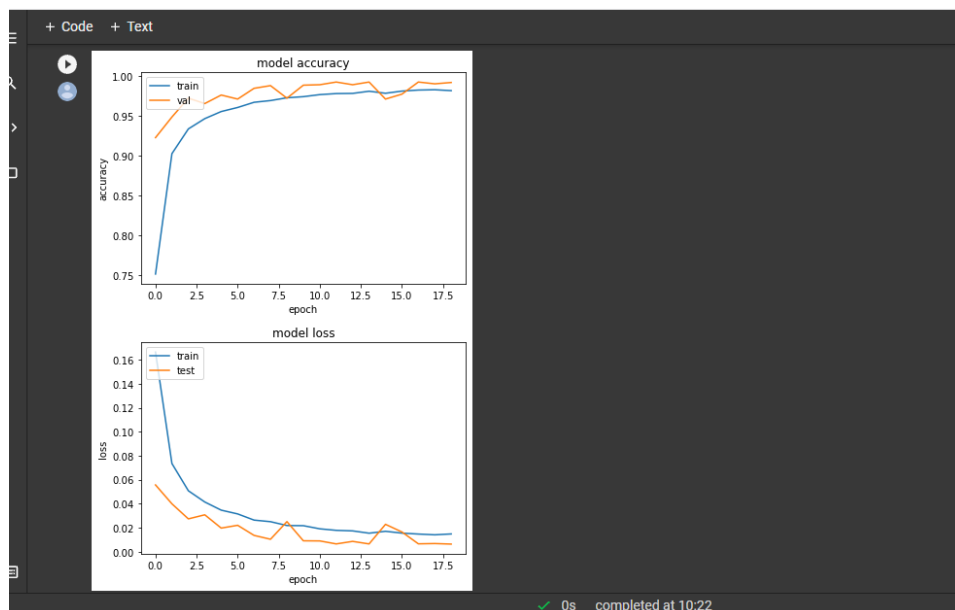


Figure 2.10 Accuracy and Loss rate

```
In [9]: # Split train and validation dataset with 10%
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.1, random_state=63)

# Show messages
print('X_train shape:', X_train.shape)
print(X_train.shape[0], 'train samples')
print(X_test.shape[0], 'test samples')

X_train shape: (15984, 64, 64, 3)
15984 train samples
1776 test samples
```

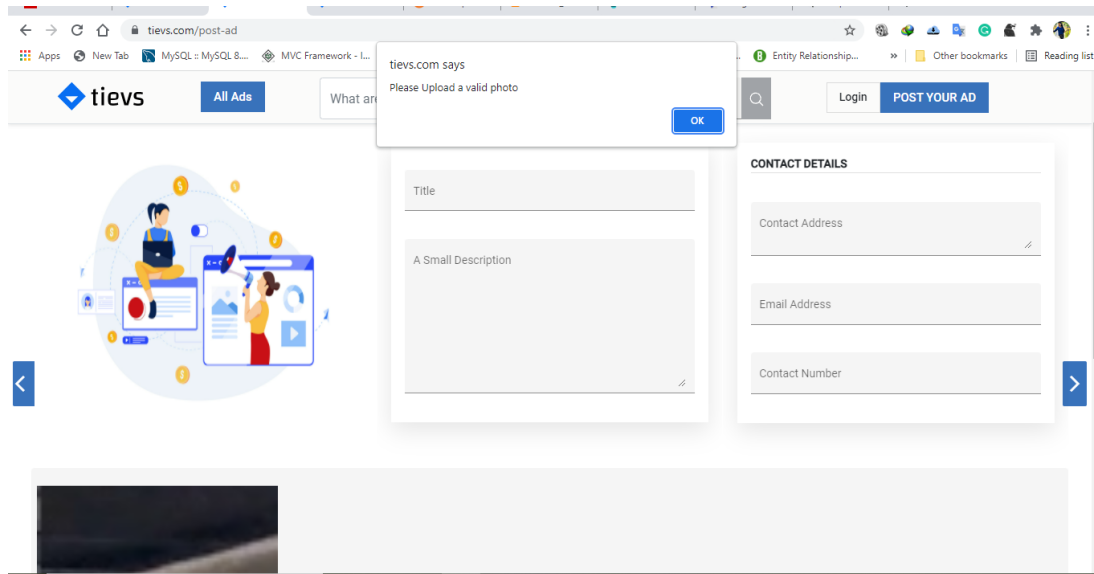
*Fig 2.11. Train the model as x and y*

```
In [15]: history = model.fit(X_train, Y_train, batch_size=BATCH_SIZE, epochs=EPOCHS, verbose=1, validation_data=(X_test, Y_test))

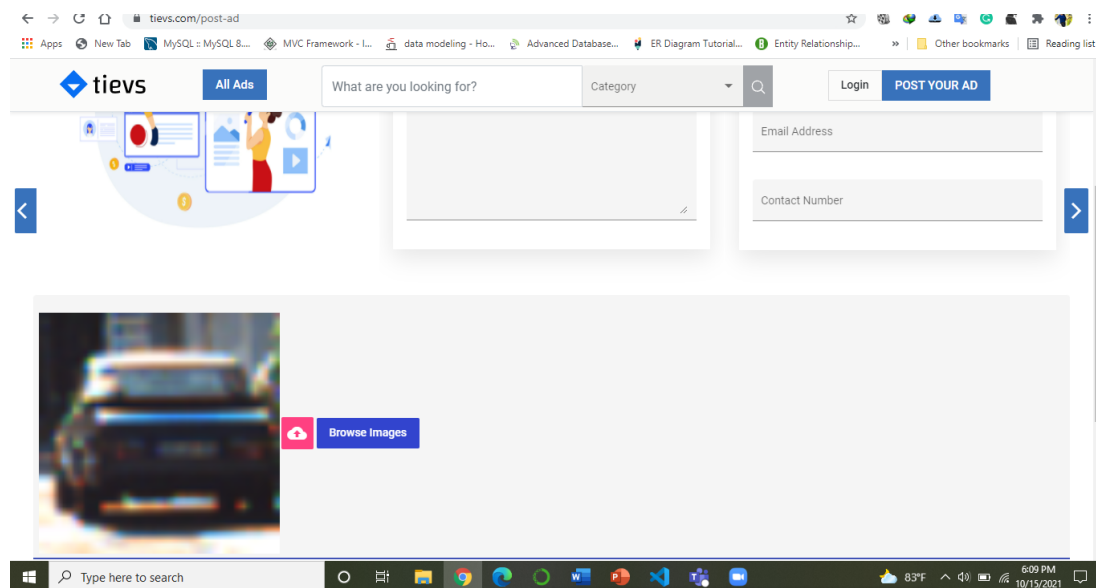
Epoch 1/20
138/138 [=====] - 45s 206ms/step - loss: 0.2358 - accuracy: 0.6333 - val_loss: 0.0556 - val_accuracy: 0.9223
Epoch 2/20
138/138 [=====] - 26s 190ms/step - loss: 0.0870 - accuracy: 0.8809 - val_loss: 0.0399 - val_accuracy: 0.9482
Epoch 3/20
138/138 [=====] - 26s 189ms/step - loss: 0.0514 - accuracy: 0.9325 - val_loss: 0.0273 - val_accuracy: 0.9718
Epoch 4/20
138/138 [=====] - 26s 187ms/step - loss: 0.0426 - accuracy: 0.9439 - val_loss: 0.0307 - val_accuracy: 0.9651
Epoch 5/20
138/138 [=====] - 26s 187ms/step - loss: 0.0351 - accuracy: 0.9533 - val_loss: 0.0196 - val_accuracy: 0.9758
Epoch 6/20
138/138 [=====] - 26s 189ms/step - loss: 0.0314 - accuracy: 0.9611 - val_loss: 0.0220 - val_accuracy: 0.9707
Epoch 7/20
138/138 [=====] - 26s 188ms/step - loss: 0.0282 - accuracy: 0.9645 - val_loss: 0.0135 - val_accuracy: 0.9842
Epoch 8/20
138/138 [=====] - 26s 187ms/step - loss: 0.0256 - accuracy: 0.9683 - val_loss: 0.0103 - val_accuracy: 0.9876
Epoch 9/20
138/138 [=====] - 26s 190ms/step - loss: 0.0233 - accuracy: 0.9708 - val_loss: 0.0250 - val_accuracy: 0.9718
Epoch 10/20
```

*Fig 2.12 Model train as x train and y train*

UI implementation after cloud deployment.



*Fig 2.13 UI implementation after uploading non-vehicle*



*Fig 2.13 UI implementation after uploading vehicle*

### **3.2 Research Finding**

- This research, use deep learning methods to classify the two classes: vehicle and no vehicle. This technique allows the development of discriminant functions immediately on the basis of data without any prior knowledge of the feature extraction procedure. Utilize convolutional neural networks (CNN), which have obtained extremely good results in numerous computer vision applications over the last several years, including image classification, object identification, image segmentation, and many more.
- This enables us to essentially tackle two problems: the first is a lack of data for training. Second, it enables us to adjust the size of the input data in accordance with the related and unrelated, even without augmentation methods or extra geometric deformations. Furthermore, because the dataset lacks a segmented annotation, a simple and efficient approach is utilized to determine the mask of the false section based on the information included in the fake images. Model is a convolutional neural architecture.

### **3.3 Discussion**

Using verification and validation approaches, application developers must create tight program logic to identify and prevent counterfeit photos from being supplied, either accidentally or intentionally. If those advertising were submitted, shoppers may eventually see incorrect ads and lose confidence in utilizing the app for their purposes, causing the platform's reputation to suffer. The above are not prioritized in current applications, nor are they monitored or prevented. Furthermore, many installed classified advertising systems do not have rich User Interfaces (UIs) for seamless functionality or to encourage a positive user experience, despite the fact that this is in line with the most recent scientific discoveries. After analyzing the issues, the authors investigated and analyzed several scientific approaches before proposing a novel solution in the form of an intelligent and advanced web application that incorporates Machine Learning (ML) and Deep Learning (DL) technology, with the primary goal of outperforming current competitors and providing customers with the best possible user experience when browsing online classifieds. As previously stated, the optimal strategy for detecting fraud images for car classifieds using deep learning appliances has been proven.

#### **4. CONCLUSION**

In comparison to magazines, and broadcast communications networks, classified advertisement interactions have grown in popularity. Using verification and validation procedures to detect and prevent the publication of deceptive advertising to the fake images, whether mistakenly or intentionally, is a huge endeavor, especially given the popularity of classifieds sites as noted in the study question above. Thousands of dollars have been lost as a result of fraudulent classified advertising, websites posing as reliable sources of information, commodities, products, and services. Multiple research and techniques have been established to detect counterfeit web classifieds applications due to the numerous negative consequences of online deception and fraudulent activities; however, none of them have been able to provide adequate and appropriate responses to suppressing these fraudulent activities. From the standpoint of an application developer, using acceptance testing methodologies, strict program logic must be written to detect and prevent fake images from being published, either intentionally or accidentally. Consumers may lose trust in the application if those photographs are published, as they may eventually view false images.

Using the CNN model successfully identify counterfeit images by paying attention to car image attributes and expressly certifying image legitimacy to prevent hoaxes. furthermore, the authors intend to determine the color of car images as well as their uniqueness prior to submission. nonetheless, this classified advertising web program was enhanced with advantageous features to ensure that online advertising stature was elevated and that competitive advantages were gained over other systems. authors will devote themselves to expanding the services now coordinated by embracing various sorts of classifieds such as online classifieds as a future channel for the overall system.

## 5. REFERENCE

- [1] Y. Guo, X. Cao, W. Zhang, and R. Wang, "Fake colorized image detection," *IEEE trans. inf. forensics secur.*, vol. 13, no. 8, pp. 1932–1944, 2018.
- [2] B. Liu, C.-M. Pun, and X.-C. Yuan, "Digital image forgery detection using JPEG features and local noise discrepancies," *ScientificWorldJournal*, vol. 2014, p. 230425, 2014.
- [3] J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz, "Leveraging frequency analysis for deep fake image recognition," *arXiv [cs.CV]*, pp. 3247–3258, 2020.
- [4] Researchgate.net.[Online].Available:[https://www.researchgate.net/publication/259950220\\_Blind\\_Fake\\_Image\\_detection](https://www.researchgate.net/publication/259950220_Blind_Fake_Image_detection). [Accessed: 23-Mar-2021].
- [5] Researchgate.net.[Online].Available:[https://www.researchgate.net/publication/342088036\\_Exposing\\_Fake\\_Images\\_With\\_Forensic\\_Similarity\\_Graphs/citations](https://www.researchgate.net/publication/342088036_Exposing_Fake_Images_With_Forensic_Similarity_Graphs/citations). [Accessed: 23-Mar-2021].
- [6] M. Laliberte, "The 9 biggest scams to avoid when buying a car on Craigslist," *Business Insider*, 25-Apr-2019.
- [7] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.
- [8] J. Xingteng, W. Xuan, and D. Zhe, "Image matching method based on improved SURF algorithm," in *2015 IEEE International Conference on Computer and Communications (ICCC)*, 2015, pp. 142–145.
- [9] "45 fake car ads from craigslist," *Cockeyed.com*. [Online]. Available: [https://cockeyed.com/citizen/accord/100\\_postings/45\\_fakes.php](https://cockeyed.com/citizen/accord/100_postings/45_fakes.php). [Accessed: 23-Mar-2021].
- [10] A. Kunbaz, S. Saghir, M. Arar, and E. B. Sonmez, "Fake image detection using DCT and local binary pattern," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2019, pp. 1–6.
- [11] "Home," *Govmu.org*. [Online]. Available: <https://nlta.govmu.org/Pages/default.aspx>. [Accessed: 23-Mar-2021].
- [1 2] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, "Deep fake image detection based on pairwise learning," *Appl. Sci. (Basel)*, vol. 10, no. 1, p. 370, 2020. [1 3] M. Ahmed, "False image injection prevention using iChain," *Appl. Sci. (Basel)*, vol. 9, no. 20, p.



4328, 2019

**6. GLOSSARY**

**7. APPENDICES**