INSURANCE AMOUNT PREDICTION BASED ON ACCIDENTAL CAR DAMAGE LEVEL USING CNN

Beecha Aditya[1], Mohammad Moizuddin[2], Mekala Naveen[3], A. Amulya[4]

[1]*Dept. of information technology, Mahatma Gandhi institute of technology*
[2]*Dept. of information technology, Mahatma Gandhi institute of technology*
[3]*Dept. of information technology, Mahatma Gandhi institute of technology*
[4]*Dept. of information technology, Mahatma Gandhi institute of technology*

---***---

Abstract: There is a paradigm shift in the automotive industry with the advancement of technology, especially in the area of insurance claims handling. In this study, we propose a new method for predicting insurance claims based on car damage levels, utilizing advanced artificial intelligence algorithms for image analysis, particularly employing Convolutional Neural Networks (CNN) for image processing and classification.

Our strategy is to provide car damage estimates through picture analysis, making insurance claims in the automobile sector simpler. In order to facilitate user interaction with Flask and increase the predictability of our models, we have also implemented enhancements to the user interface.

By combining image analysis with a user-friendly interface, our method offers a practical solution for insurance companies to expedite claim processing, reduce fraudulent activities, and ensure fair settlements for policyholders. This research contributes to the ongoing transformation of insurance claim processing within the automotive industry, highlighting the potential of merging image analysis methodologies with user interfaces for enhanced efficiency and usability.

Keywords: Insurance claim prediction, Convolutional Neural Networks (CNNs), Image analysis, User interface, Flask, Automotive industry.

1.INTRODUCTION

In the domain of car insurance, the issue of claims leakage, defined as the disparity between the actual claim payment and the rightful amount, poses a significant challenge. Traditional approaches to mitigate this problem have predominantly relied on manual validation and visual inspection processes, which are time-consuming and prone to errors. While some startups have attempted to address these inefficiencies by introducing faster claim processing methods, there remains a need for more innovative solutions.

Our project aims to tackle this challenge by introducing a novel approach to predict insurance amounts based on the level of accidental car damage and the brand of the car. Unlike existing methods, which often rely on subjective assessments, we leverage advanced Convolutional Neural Network (CNN) methodologies to automate and enhance the accuracy of damage prediction and brand recognition.

The classification tasks of our project are twofold. First, we aim to predict the extent of damage in car accidents, from minor to severe, and identify normal damage-free conditions. This task evaluates different levels of damage, such as mild, moderate, severe and normal. In addition, we aim to identify the car brand by distinguishing visible features and design elements specific to certain brands such as Audi, BMW and Toyota.

The complexity of these tasks arises from factors such as variations in lighting conditions, camera angles, and the presence of barely visible damages. To address these challenges, we meticulously curated a comprehensive dataset by collecting diverse images from various sources and manually annotating them for damage severity and brand information.

By utilizing advanced machine learning methodologies, specifically focusing on Convolutional Neural Networks (CNNs), we aim to develop accurate and reliable models for damage prediction and brand recognition. By leveraging CNN methodologies, we seek to overcome the inherent challenges associated with image classification tasks, ultimately facilitating more efficient insurance claim processing and offering important viewpoints for insurance firms and policyholders.

Problem Statement
In the insurance industry, accurately determining compensation for car damage resulting from accidents is a challenge due to imprecise assessment methods, manual inspections causing delays and disputes, and the absence of a standardized, technologically advanced system.

This project addresses these shortcomings, aiming to streamline claims processing, reduce fraud, and enhance customer satisfaction by developing a more efficient and equitable approach to predicting insurance amounts based on accidental car damage levels.

Existing system
The existing system integrates machine learning methodologies within auto insurance operations. Its objective is to effectively manage the rising frequency and severity of auto insurance claims. The study investigates the applicability of machine learning models in predicting claim incidents and adjusting insurance rates accordingly. The dataset utilized in this research is supplied by Seguro, a prominent automotive insurance company based in Brazil. A variety of machine learning algorithms, including K-NN, naive Bayes, random forest, decision trees, XGBoost and logistic regression, are deployed and assessed for their performance. The findings indicate that the random forest model outperforms others, demonstrating high accuracy, kappa, and AUC values. This initiative underscores the potential of machine learning in enhancing the understanding and utilization of large-scale insurance data to provide more effective customer service in the car motors insurance sector.

However, the system heavily relies on manual data entry and processing, leading to errors and inefficiencies, Traditional methods lack precision in predicting risk factors and insurance premiums due to their reliance on historical data, these systems struggle to adapt to changing market conditions, emerging risks, and evolving customer behaviours, The manual processes and lack of automation result in elevated operational expenses for insurance companies.

2.LITERATURE SURVEY

Ruixing Ming and Mohamed Hanafy [1] explored machine learning algorithms for auto insurance, emphasizing the effectiveness of random forest algorithm model in predicting claim occurrence for efficient customer service.

Abdelhadi et al. [2] suggested a framework for forecasting auto insurance claims, emphasizing the significant challenges posed by limited training data for accurate accident prediction. They underscored the significance of tackling these challenges to enhance the dependability of insurance claim prediction models.

Qinghui Zhang et al. [3] developed an advanced vehicle damage detection algorithm, highly precise, built on the Mask RCNN framework. Their algorithm utilized advanced techniques such as region proposal network (RPN) and fully convolutional network (FCN), achieving remarkable performance in accurately identifying and isolating affected areas in automobile images.

R. Girshick et al. [4] introduced the R-CNN technique for object detection. Their approach, which includes region proposals, CNN feature extraction, and class-specific linear SVMs, demonstrated promising results in effectively classifying region proposals based on extracted features, showcasing its potential for various object detection tasks.

Ahmad Bahaa Ahmad et al. [5] explored machine learning for vehicle classification via seismic fingerprinting. Logistic Regression (LR), Support Vector Machine (SVM), and Naïve Bayes (NB) algorithms were used. The study compared these methods with a CNN architecture, assessing precision, recall, f1-score, and accuracy metrics. It aimed to evaluate LR, SVM, and NB for vehicle classification based on seismic signals, contrasting them with CNN.

Yi Ouyang et al. [6] presented an algorithm for robust vehicle logo detection and classification using coarse-to-fine localization, multi-class structural SVM, spike adjustments for sparse coding, feature selection, and multi-feature learning. Their work was presented at the 5th International Conference focusing on Measurement, Instrumentation and Automation (ICMIA 2016).

Jessica Pesantez-Narvaez, Montserrat Guillen, Manuela Alcañiz [7] compared XGBoost and logistic regression for motor insurance claims prediction with telematics data. XGBoost exhibited higher accuracy but needed careful regularization to prevent overfitting. Logistic regression offered straightforward interpretation.

Subsequent studies may explore expanded datasets and alternative machine learning approaches.

Israfil Ansari, Yeunghak Lee, Yunju Jeong, Jaechang Shim [8] presented a practical system for vehicle logo detection and recognition, leveraging the faster R-CNN model and perspective transformation. Their work, titled "Recognition of Car Manufacturers using Faster R-CNN and Perspective Transformation," addresses challenges like complexity and low detection rates, particularly from street top view cameras. The system efficiently detects and labels vehicle logos, enhancing traffic monitoring systems' intelligence. Perspective transformation significantly improves results compared to normal CCTV images, although night data quality remains a challenge. Future work aims to enhance night image quality for improved logo detection and recognition.

Siddhant Gole, Pranay Gupta, Gauri Sanjay Patil, and Padmashri Vijayavel [9] present a system titled "Car Damage Assessment to Automate Insurance Claim," aiming to detect and classify damaged parts of car images. The system utilizes image processing, deep learning, and transfer learning techniques to generate a cost report for damaged cars. Future improvements focus on enhancing accuracy in recognizing number plates and damaged parts.

3.METHODOLOGY

Proposed system

The proposed system, involves the implementation of a user-friendly website equipped with Convolutional Neural Network (CNN) technology. This platform allows users to upload images of damaged cars, enabling the system to perform a sophisticated analysis of the damage severity. Leveraging CNN's image recognition capabilities, the system then provides accurate predictions for the insurance amount associated with the car damage. This web-based solution aims to streamline the insurance claims process by eliminating manual assessments, reducing the risk of fraudulent claims, and enhancing efficiency. Through the integration of CNN and image uploading functionality, the system provides a user-centric approach to predicting insurance amounts based on the level of accidental car damage.

Convolutional Neural Networks (CNNs) offer substantial benefits, notably enhancing prediction accuracy by autonomously extracting complex features from car damage images, enabling more precise claim amount estimations, CNNs demand less manual feature engineering and upkeep in contrast to conventional machine learning models. Deep learning models possess the capability to continuously learn and adjust to novel data, ensuring the system remains up-to-date and capable of handling evolving trends in car damage assessment.

System Architecture

The system architecture for "Insurance Amount Prediction Based on Accidental Car Damage Level" comprises two distinct tasks aimed at efficient insurance claim processing in the automotive sector. The first task involves damage level classification, where car damage images undergo preprocessing and feature extraction to classify them into different damage levels, such as minor, moderate, severe, and normal. These images are initially processed to enhance quality and diversity, followed by the extraction of relevant features. These features serve as input to a classification model, including Convolutional Neural Networks (CNNs) or custom CNN models like ResNet50, trained to accurately categorize the images depending on the severity level of the damage.

In the second task, the focus shifts to car brand recognition, wherein damaged vehicle images undergo similar preprocessing and feature extraction steps. However, in addition to damage-related features, brand-related features are also extracted, capturing unique characteristics specific to different car brands, such as logos or distinctive design elements. Using these features, another classification model, such as a CNN model, is trained to recognize the brand of the damaged vehicle. These two tasks operate independently but contribute synergistically to the overarching goal of predicting insurance amounts accurately. By combining damage level classification and car brand recognition, the system aims to provide comprehensive information for streamlined insurance claim processing, facilitating faster and more efficient decision-making in the insurance industry.

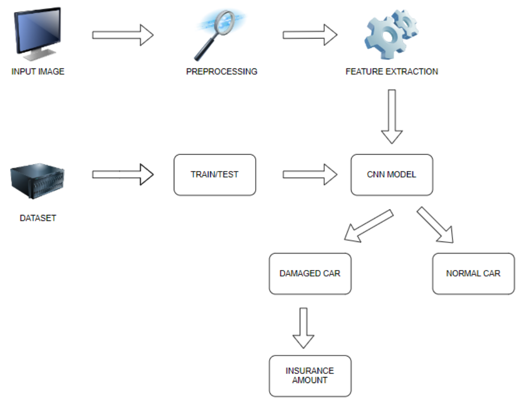


Fig.1. Proposed system

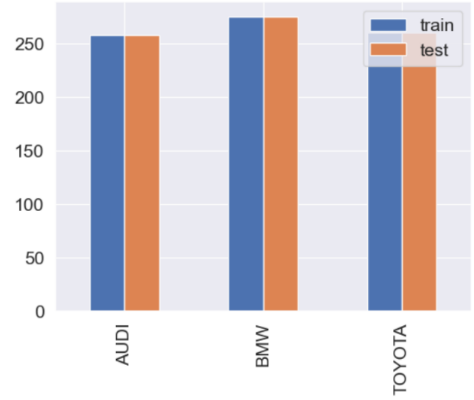


Fig.3. Brand dataset

Dataset

The dataset for "Insurance Amount Prediction Based on Accidental Car Damage Level" comprises two distinct sets of images. In the first phase, the dataset focuses on categorizing car damage into different severity levels, including minor, moderate, severe, and normal conditions. These classes enable binary classification initially, distinguishing between normal and significant damage levels.

Furthermore, an independent dataset was analyzed for brand recognition, including of pictures representing cars from different manufacturers, such as Audi, BMW, and Toyota. This brand database makes it possible to categorize vehicles according to their brand, which aids in the evaluation of insurance benefits overall.

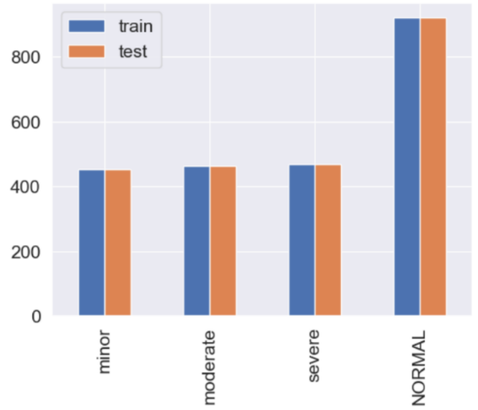


Fig.2. Damage level dataset

Image Processing

Image processing is pivotal in the initial preprocessing stage of car damage classification using images. Raw images of damaged cars undergo various preprocessing steps to enhance quality and extract pertinent features. Techniques such as noise reduction, contrast enhancement, and normalization are applied to ensure consistency and improve image clarity. Additionally, image registration techniques may be utilized to align images, facilitating accurate comparison and analysis. Segmentation methods are then employed to isolate regions of interest, such as damaged areas, from the background, enabling the extraction of relevant features for subsequent analysis. These preprocessing techniques optimize image quality, consistency, and interpretability, laying a solid groundwork for subsequent model development and classification.

Algorithms

Convolutional Neural Networks (CNNs) is selected for their capacity to extract features from images and recognize patterns effectively, making them ideal for tasks such as object detection and image classification in the context of car damage assessment.

ResNet50: ResNet50, short for Residual Neural Network with 50 layers, is a deep learning architecture designed to address training difficulties encountered in very deep neural networks. It employs residual connections, allowing for smoother information flow during training and mitigating the vanishing gradient problem

Custom CNN Model: A custom CNN model is developed specifically for car damage classification. This model comprises convolutional layers deployed for feature extraction, followed by pooling layers for dimensionality reduction and dense layers for classification. By training this model on the dataset, it learns to accurately classify car damage images into different severity levels, contributing to efficient insurance claim processing.

Modules

1. **Input image acquisition**: The process of obtaining an image from a source, typically hardware equipment like cameras, sensors, etc. is known as input image acquisition in image processing and machine vision.
2. **Preprocessing**: To improve the quality and usefulness of machine learning models, preprocessing refers to the processes done to format the images before they are utilized for model training and inferences.
3. **Feature Extraction**: This involves a normalising feature scale to allow a fair comparison, correcting missing error values and selecting the most important features and features from the raw data.
4. **Feature analysis**: This is a crucial step in understanding the relevance and impact of individual features on a given dataset, helping to uncover patterns and relationships within the data. By examining the statistical properties and distributions of features, we can gain insights into their contribution to the overall problem.

Applications

- **Reduction in Human Error:** Automated accident prediction technology minimises the human error by providing objective and consistent estimates, reducing the likelihood of biased estimates.
- **Zero Risk:** Although achieving absolute zero risk is almost impossible, strict safety measures must be implemented and advanced technologies such as predictive damage analysis of car accidents must be used.
- **24X7 Availability:** As we are providing a website for the prediction of insurance amount, we promise 24X7 availability of the website.
- **Digital Assistance:** Digital assistance revolutionizes daily operations by providing instant guidance and problem solving with the help of artificial intelligence.
- **Unbiased Decision:** Unbiased decision-making, facilitated by objective algorithms and data-driven analysis, ensures fairness and impartiality in the evaluation process.
- **Prevention of Fraudulent Claims:** Automated damage predictions act as a deterrent to fraudulent claims, reducing the occurrence of errors associated with false or exaggerated damage reports.
- **Emergency Response Planning:** Emergency response agencies can use predictive models to anticipate the severity and impact of car accidents in specific regions
- **Smart City Initiatives:** Municipalities and urban planners can integrate predictive models into smart city initiatives to enhance transportation infrastructure and traffic management systems. By analyzing historical accident data and real-time traffic patterns, cities can optimize road networks, improve traffic flow, and reduce the risk of accidents.

4.RESULTS

In damage level classification, the model showed a high accuracy of 0.8976, indicating its ability to accurately classify car damage into minor, moderate, severe and normal conditions. This level of accuracy is critical to insurance appraisals because it provides an accurate estimate of repair costs and benefits based on the severity of the damage.

Similarly, in the brand recognition task, the model achieved a commendable accuracy of 0.8793 in correctly identifying the make of cars from images. This capability is essential for insurance companies to determine the value and replacement costs of vehicles involved in accidents accurately.

Fig.4. Home Page

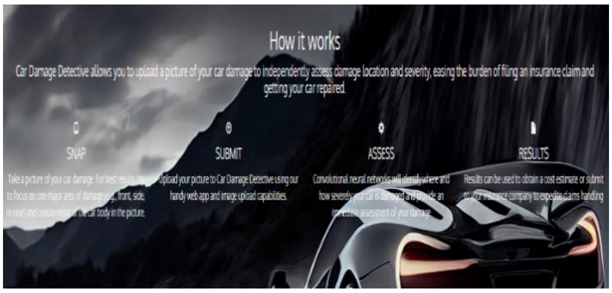


Fig.5. how it works

Fig.6. upload image tips

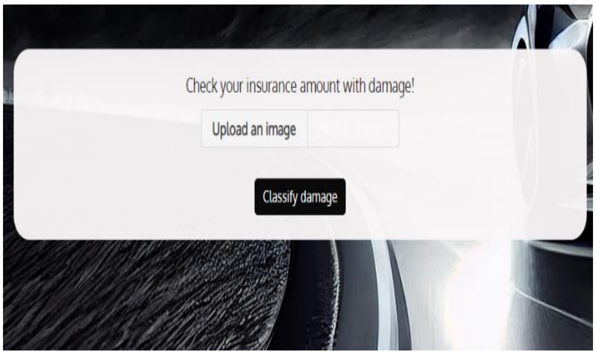


Fig.7. image upload area for insurance amount with damage level

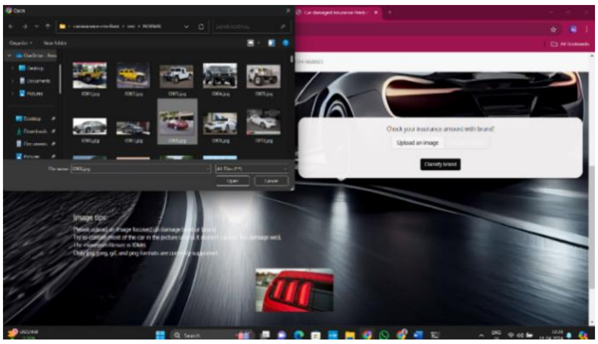


Fig.8. uploading image

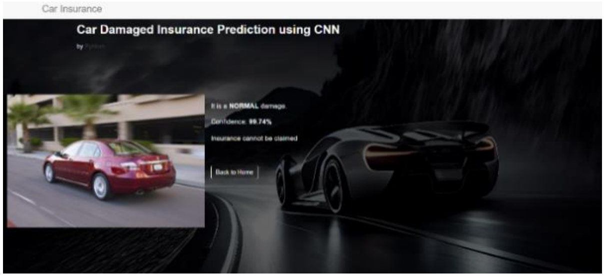
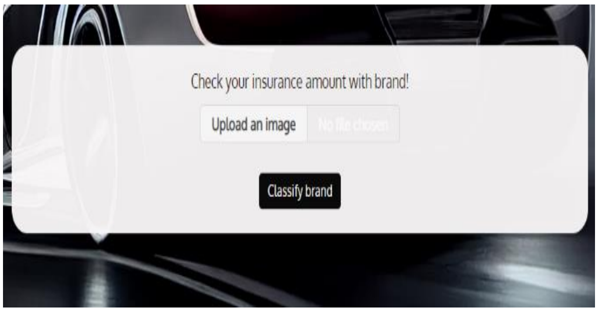


Fig. 9 output



Fig,14. image upload area for insurance amount with brand

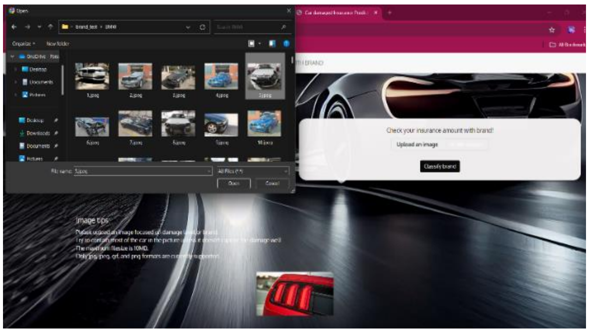
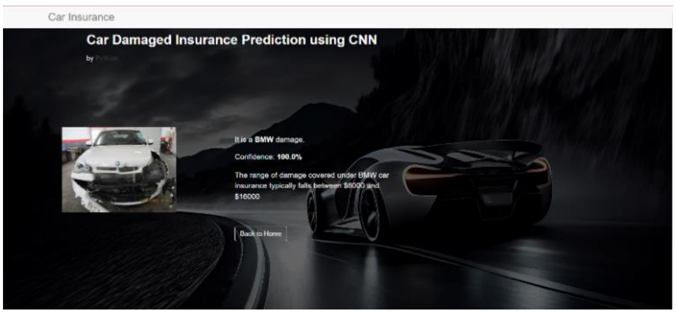


Fig.15. uploading brand image



Fig,16. final output

5.CONCULSION

Using advanced machine learning techniques, specifically Convolutional Neural Networks (CNN), the project successfully meets the challenge of accurately predicting insurance amounts based on the severity of car claims and the make of the vehicle in question.

Through the development and evaluation of models for damage level classification and brand recognition, the project demonstrates promising results. With a high accuracy of 89% in damage level classification and 87% in brand recognition, the models showcase their effectiveness in accurately assessing the extent of damage and identifying the make of cars from images, respectively.

These results highlight the potential of the proposed approach to streamline the insurance claim assessment process, resulting in a more efficient and accurate determination of loss amounts. By automating this critical part of the insurance industry, the project aims to improve customer service, reduce processing times and improve the overall operational efficiency of insurance providers.

Future research and development could focus on refining the models, expanding the dataset to include more diverse samples, and incorporating additional features to improve predictive accuracy. In addition, actual implementation and integration with existing insurance systems could be investigated to assess the scalability and practicality of the proposed solution in a production environment.

6.FUTURE SCOPE

Moving forward, there are numerous avenues for advancing insurance claim prediction based on car damage assessment. Enhancing model robustness and accuracy can be achieved by diversifying datasets to include a broader range of damage types, severity levels, and vehicle brands. This expanded dataset would bolster the models' ability to adapt to various scenarios.

Further exploration into advanced machine learning techniques like ensemble learning, transfer learning, and attention mechanisms could significantly enhance predictive performance. Additionally, fine-tuning hyperparameters and optimizing model architectures hold promise for achieving better results.

Integrating additional data sources such as telematics, accident reports and weather conditions can provide valuable context for more accurate forecasts. In addition, the use of new technologies such as computer vision, natural language processing and blockchain can revolutionize insurance claims by taking care of efficiency, transparency and security.

Improving the user interface and experience could enhance adoption among insurance professionals and policyholders. Intuitive visualization tools, interactive dashboards, and real-time updates would streamline claim assessment processes and increase satisfaction.

Lastly, exploring the broader application of predictive models beyond insurance claim prediction could uncover new opportunities. These models could find utility in vehicle maintenance scheduling, risk assessment for underwriting purposes, or fraud detection in insurance claims, further expanding their impact and relevance.

REFERENCES

[1] Mohamed Hanafy, Ruixing Ming,” ML methods for Auto Insurance Big Data”, MDPI,2021.

[2] Abdelhadi, Shady, Khaled Elbahnasy, and Mohamed Abdelsalam 2020. “A model suggested for forecasting auto insurance claims using ML techniques.” Journal of Theoretical and Applied Information Technology 98: 3428–3437.

[3] Quinghui Zhang, Xianing Chang, Shanfeng Bian Bian, “Vehicle-Damage and Detection segmenting technique based on RCNN” IEEE international conference,2019, DOI 10.1109/ACCESS.2020.2964055.

[4] T. Darell, R. Girshick, J. Donahue, et al. , ”Rich feature hierarchies for accurate object detection and semantic segmentation. ”IEEE conference on computer pattern and vision recognition,Jan.2014,vol.13, no.1, pp. 580–587.

[5] Ahmad, A. B., Saibi, H., Belkacem, A. N., & Tsuji, T. (2022). Vehicle Classification with the help of ML techniques Based on Seismic Fingerprinting. Computers, 11, 148.

[6] Y. Ouyang, “Vehicle Logo Detection and Classification using Discriminative pixel-patches Sparse Coding,” in 5th International

Conference on Measurement, Instrumentation and Automation, 2016.

[7] Jessica Pesantez-Narvaez, Montserrat Guillen and Manuela Alcañiz, “Predicting the car Insurance Claims with help of Telematics Data LR versus XGBoost”, MDPI, Risks 2019, vol. 7, 70; doi:10.3390/risks7020070

[8] I. Ansari, Y. Lee, Y. Jeong, and J. Shim, “Recognition of Car Manufacturers using Faster R-CNN and Perspective Transformation,” Journal of Korea Multimedia Society, vol. 21, no. 08, 2018.

[9] Siddhant Gole, Pranay Gupta, Gauri Sanjay Patil and Padmashri Vijayavel, “Car Damage Assessment to Automate Insurance Claim”, International Research Journal of Engineering and Technology (IRJET), Volume: 09 Issue: 03, Mar 2022.

[10] Ross Girshick, Piotr Dollár, Georgia Gkioxari and Kaiming He, "Mask r-cnn", Proceedingsof the IEEE international conference on computer vision, pp. 2961-2969, 2017.

[11] [online] Available: https://www.kaggle.com/datasets/prajwalbhamere/car-damage-severity-dataset.