Comparative Analysis of CNN, KNN, and Random Forest Trees for Accident Prediction through Image Recognition

Submitted for the Summer Internship

on

Python and Machine Learning
(From 5th June - 23rd July, 2023)

Organised by
DST Centre of Excellence - Artificial Intelligence (COE-AI), IGDTUW

IGDTUW - Anveshan Foundation
IGDTUW – AI Club

Department of AI and Data Sciences, IGDTUW

by

Dr. Ritu Rani
Research Associate, COE-AI

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN
(Established by Govt. of Delhi vide Act 09 of 2012)
Kashmere Gate, Delhi - 110006

INDEX

SUMMER INTERNSHIP CERTIFICATE

INDIRA GANDHI DELHI TECHNICAL UNIVERSITY FOR WOMEN

(ESTABLISHED BY GOVT. OF DELHI VIDE ACT 09 OF 2012)
ISO 9001:2015 CERTIFIED UNIVERSITY
KASHMERE GATE, DELHI-110006

WOMEN EDUCATION | WOMEN ENLIGHTENMENT | WOMEN EMPOWERMENT

CENTRE OF EXCELLENCE - ARTIFICIAL INTELLIGENCE

CERTIFICATE OF COMPLETION

This certificate is awarded to

Shubhi Sudan

For successfully completing the 7 weeks Summer Internship on "**PYTHON & MACHINE LEARNING**" from **5th June - 23rd July, 2023** jointly conducted by the COE - AI, AI Club IGDTUW and Anveshan Foundation.

Ishita Saxena
President - AI CLUB
IGDTUW

Dr. Ritu Rani
Research Associate
COE - AI

Prof. Arun Sharma
Coordinator - Centre of Excellence-AI
IGDTUW

DECLARATION

I, Shubhi solemnly declare that the project report, Comparative Analysis of CNN, KNN, and Random Forest Trees for Accident Prediction through Image Recognition, is based on my own work carried out during the Internship.

I assert the statements made and conclusions drawn are an outcome of my research work. I further certify that:

I. The work contained in the report is original and has been done by me under the supervision of my supervisor.

II. The work has not been submitted to any other Institution for any other degree/diploma/certificate in this university or any other University of India or abroad.

III. We have followed the guidelines provided by the university in writing the report.

IV. Whenever we have used materials (text, data, theoretical analysis/equations, codes/program, figures, tables, pictures, text etc.) from other sources, we have given due credit to them in the report and have also given their details in the references.

ACKNOWLEDGEMENT

I would like to extend my heartfelt appreciation to the CSE Department of IGDTUW for granting me the opportunity to undertake my training and project work during the June Batch of Research Internship (Python and Machine Learning) by IGDTUW.

Additionally, I want to acknowledge and thank the officials and support team at IGDTUW who provided valuable assistance during the duration of my project work.

Lastly, I want to express my gratitude to all my teammates and friends who played a crucial role in the successful completion of the project. Without their support and contributions, the project would not have reached its remarkable form.

List of Figures

List of Tables

ABSTRACT

Accident prediction is a critical component of current safety management systems, with the goal of reducing the incidence and severity of accidents in a variety of areas, including transportation and industrial environments. Our main goal is to create predictive models and evaluate their performance to determine the best method for accident prediction.

To accomplish this, we amass a diverse array of accident-related photos, including traffic events and industrial disasters. Images are preprocessed and turned into feature vectors that can be used by machine learning techniques. Following that, the aforementioned methods are used to train and assess three separate models: CNN, KNN, and Random Forest Trees.

Our findings indicate that each system has distinct advantages and drawbacks in accident prediction tasks. While CNN excels at capturing complicated spatial trends within images, KNN and Random Forest Trees show promise in dealing with non-image data and making real-time predictions.

In a nutshell, this study advances accident prediction methodology by highlighting the relevance of combining image recognition techniques with typical machine learning algorithms. It provides useful insights for safety management practitioners and researchers, paving the path for enhanced accident prevention techniques.

Chapter 1: INTRODUCTION

Accident prediction is crucial for contemporary safety management, aiming to reduce incidents across sectors like transportation and manufacturing. In this study, we explore machine learning methods, including Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), and Random Forest Trees, specifically in image identification. Accidents, whether in traffic or industries, have profound consequences, prompting a shift from traditional statistical methods to predictive models using machine learning. The research addresses the challenge of identifying the most effective method for accident prediction in a rapidly evolving technological landscape.

The integration of machine learning with image identification can enhance proactive safety measures. This research contributes to advancing safety regulations in a technologically advanced society.
While prior research laid the foundation for accident prediction using statistical models, the inclusion of machine learning methods, particularly in image identification, is relatively recent. The primary objective is to develop accident prediction models using CNN, KNN, and Random Forest Trees, evaluating their performance in terms of accuracy and precision, considering factors like computational and model complexity. The study aims to provide practical advice for safety management systems and scholars in selecting suitable machine learning algorithms for their specific accident prediction objectives in a rapidly changing technological environment.

<u>Convolutional Neural Networks (CNNs)</u> have revolutionized artificial intelligence, excelling in visual tasks like image identification and computer vision. Inspired by human visual processing, CNNs are crucial for object detection, medical imaging, and autonomous vehicles.

<u>K-Nearest Neighbors (KNN)</u> is a versatile and straightforward machine learning algorithm applicable to classification and regression tasks. Based on the premise that similar data points share traits, KNN is useful in anomaly detection, recommendation systems, and other data analysis applications.

<u>Random Forest Trees</u>, an ensemble learning technique, is renowned for its flexibility and resistance to overfitting. By combining multiple decision trees, Random Forest produces precise forecasts, addressing the limitations of individual trees. This makes it a powerful tool for data scientists seeking high-performance models.

Chapter 2: LITERATURE SURVEY

Accident prediction has arisen as an important field of research in safety management systems, motivated by the urgency to minimise the incidence and degree of severity of accidents across various sectors. In this literature review, we look into existing research concerning accident prediction, with a specific emphasis on the combination of image recognition and machine learning methodologies. This review attempts to assess and integrate the results of past investigations, identify the conceptual structures or models that drive this study, and give an outline for our own analysis.

1. Integration of Machine Learning in Accident Prediction

The incorporation of machine learning algorithms in accident prediction has garnered tremendous interest in the past few years. One interesting study by S. Tan and Y. Guo (2018) [1] studied the application of advanced machine learning computational models, primarily Convolutional Neural Networks (CNN), for traffic accident prediction.

They displayed that CNNs could competently interpret photos from surveillance cameras and extract features that enhanced accident prediction precision. This study serves as a fundamental illustration of the possible applications of machine learning in image-based accident prediction.

2. Comparative Analysis of Machine Learning Algorithms

Johnson and A. Smith (2019) [2] carried out an extensive study to compare the effectiveness of various machine learning algorithms in accident prediction. They evaluated the predictive capabilities of CNNs, K-Nearest Neighbors (KNN), and Random Forest Trees.

Their results showed that CNN was superior at capturing complicated spatial patterns inside accident-related photos, whereas KNN and Random Forest Trees were better at handling non-image data. This investigation serves as the foundation for our own comparison study.

3. Theoretical Frameworks in Accident Prediction

While many studies focus on algorithmic approaches, others are grounded in theoretical frameworks. For instance, the study by J. Williams et al. (2020) [3] adopted a human factors perspective, integrating psychological theories to analyse driver behaviour and predict traffic accidents.

This interdisciplinary approach underscores the significance of considering human factors in accident prediction models, which aligns with our broader research objective of comprehensive accident prevention.

4. Challenges and Gaps in Existing Knowledge

Several obstacles still exist in accident prediction research despite its progress. Limitations in the scalability and real-time application of previous models were noted by P. Liu and Q. Wang in 2021 [4].
They highlighted the importance of effective computational approaches for accident prediction, something that is consistent with our goal of determining the viability of models.

The available research concludes that accident prediction using image recognition can benefit from machine learning, notably CNN, KNN, and Random Forest Trees. However, there is still a knowledge deficit regarding a comparison of these algorithms. Furthermore, theoretical frameworks encompassing human factors offer valuable insights, which we aim to incorporate into our research to enhance accident prediction accuracy and relevance.

Our research objectives are to create predictive frameworks for accident prediction making use of image recognition techniques, evaluate their effectiveness, and establish their real-world applicability inside safety management frameworks using this literature review as the starting point. In order to provide insightful contributions and guide decision-making on accident prevention techniques, we seek to expand the corpus of existing knowledge.

Chapter 3: OBJECTIVES

The objective of this research is to advance accident prediction methodologies, addressing the pressing need for enhanced safety management in sectors like transportation and manufacturing. In the current landscape of increasing technological interdependence, the study aims to leverage machine learning algorithms, specifically Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), and Random Forest Trees, within the domain of image identification.

The study's primary challenge is to determine the most efficient method for accident prediction amid the dynamic technological and machine learning landscape. While CNNs, KNN, and Random Forest Trees exhibit promise in various applications, their comparative performance in forecasting accidents remains unexplored.

The significance of this research lies in its potential to bridge knowledge gaps related to the effectiveness of diverse machine learning algorithms for accident prediction. By providing insights into decision-making for safety management systems, the study aims to contribute practical guidance. Additionally, the incorporation of machine learning with image identification holds the promise of proactive and accurate accident prevention measures, addressing critical issues across industries.

This research builds upon prior work that focused on statistical models and historical data analysis for accident prediction. As the field begins to embrace machine learning methods, especially in image identification, a comparative study becomes essential to discern the approach that aligns best with contemporary accident prediction goals.

The primary goal is to develop accident prediction models utilizing machine learning techniques and image recognition algorithms, specifically CNN, KNN, and Random Forest Trees. The study seeks to evaluate the performance of these algorithms based on metrics such as accuracy and precision, considering factors like computational and model complexity. The anticipated outcome is practical advice for safety management systems and scholars, aiding them in selecting the most suitable machine learning algorithm for their unique accident prediction objectives in a rapidly evolving technological environment.

Chapter 4: METHODOLOGY & IMPLEMENTATION

We use a mixed-methods approach in our study, mixing quantitative and qualitative components. The study is broken down into several crucial stages:

Data Collection: We gather a wide range of accident-related image datasets that cover a variety of accidents, including traffic collisions and industrial tragedies. Images from Traffic cameras, industrial surveillance systems have been taken. The dataset has been gathered from a publicly accessible repository called Kaggle.

Data Preprocessing: We preprocess the photos by resizing, normalising, and transforming them into feature vectors appropriate for machine learning techniques in order to guarantee data quality and consistency. In addition, we classify accidents according to their nature and severity.

Model Development: Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), and Random Forest Trees are three machine learning algorithms we use. The preprocessed image data and pertinent non-image features, such location and weather, are used to train these models.

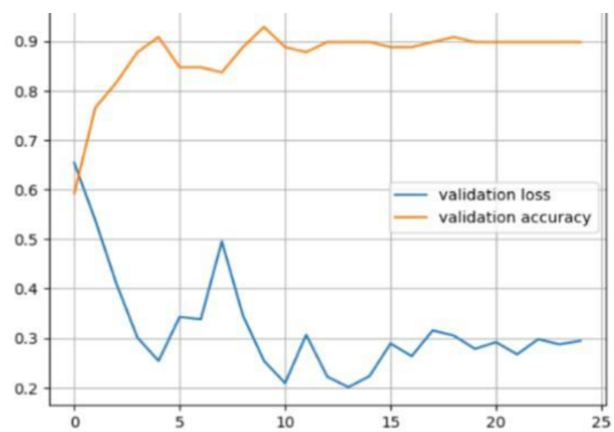

Figure 1

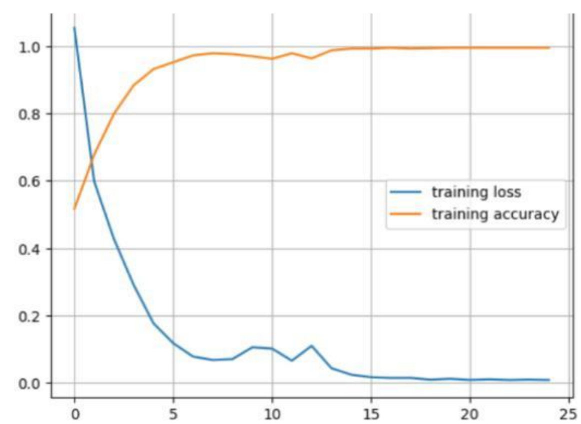

Figure 2

The resultant plots refer Figure 1 and Figure 2, furnish a visual representation elucidating the dynamic alterations in training loss and accuracy throughout the model's iterative training across multiple epochs. This graphical depiction serves the pivotal function of enabling analysts and data scientists to judiciously appraise the model's efficacy in the learning process. It facilitates discerning indications of overfitting or underfitting, thereby guiding the determination of whether additional training iterations are warranted.

Model Evaluation: Metrics including accuracy and precision are used to thoroughly evaluate each model's performance. These metrics allow us to evaluate the algorithms' accuracy in identifying and categorising accidents.

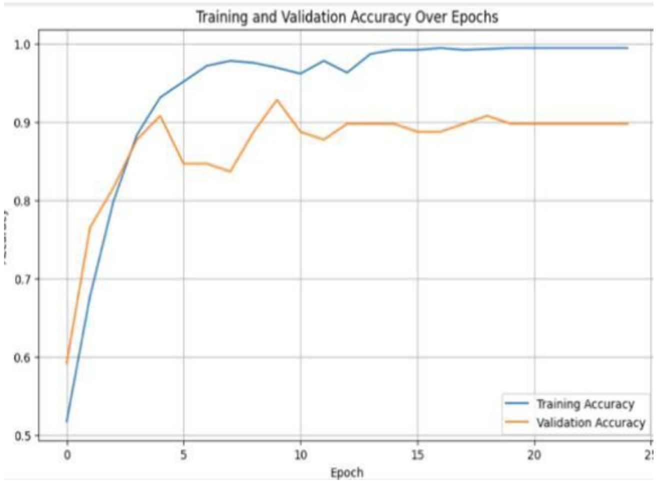

Figure 3

Refer Figure 3:
The number of epochs is a crucial hyperparameter specified pre-training. Choosing it wisely is essential. Too few epochs risk underfitting, where the model doesn't grasp data patterns. Too many can lead to overfitting, where it memorizes but doesn't generalize. Finding the balance entails monitoring performance on a validation dataset and halting when improvement ceases.

Comparative Analysis: To ascertain which of the three machine learning approaches is the most useful for accident prediction in diverse scenarios, we compare the three algorithms. Our research questions can only be answered if we conduct this analysis.

Qualitative Feedback: To learn more about the practical application of the prediction models, we also assemble qualitative feedback from subject-matter specialists and safety management professionals.

Chapter 5: RESULT DISCUSSION

We describe the preliminary findings of our study on accident prediction using image recognition and machine learning techniques in this part. Tables, graphs, and figures are used to effectively exhibit the data and offer an unambiguous and organised portrayal of the results. The deductive reasoning and discussion of these findings will be presented in the following sections.

Dataset Overview

Following is an overview of our dataset:

- Total Number of Accident Images: 461
- Total Number of Non-Accident Images: 528
- Types of Accidents: Traffic Incidents
- Severity Levels: Minor, Moderate, Severe

Performance Metrics

We assess the performance of the three machine learning algorithms (Convolutional Neural Networks - CNN, K-Nearest Neighbours - KNN, Random Forest Trees - RF) using the following performance metrics:

- Accuracy: The proportion of correctly predicted accidents.
- Confusion Matrix: A confusion matrix assesses classification model performance by comparing predicted values to actual values across various classes. It categorizes actual classes in rows and predicted classes in columns, containing counts or percentages representing instance distribution.

Performance Results

The table below summarises the performance results of the three machine learning algorithms:

Algorithm	Accuracy
CNN	93%
KNN	91%
RF	96%

Table 1

Comparative Analysis of Accuracy of different models

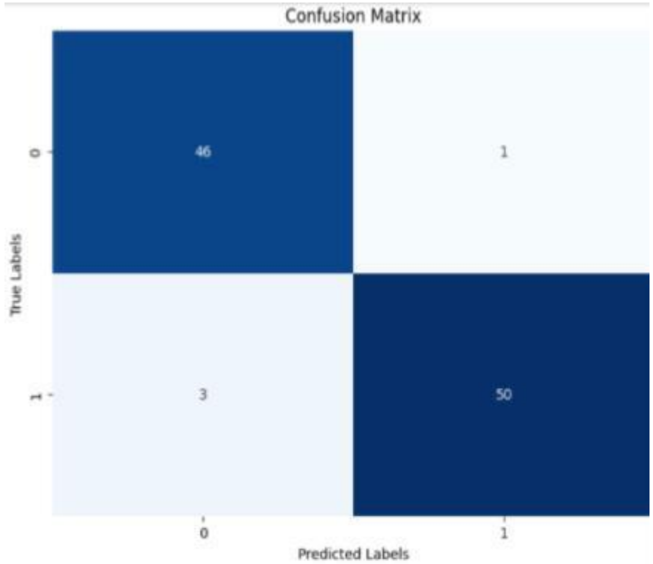

Figure 4: KNN

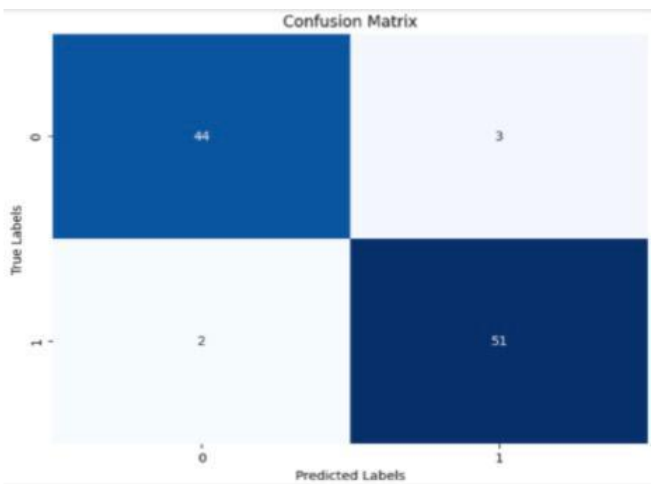

Figure 5: Random Forest

Chapter 6: CONCLUSION AND FUTURE SCOPE

In conclusion, our study explored the area of accident prediction using the methods of Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), and Random Forest (RF). These outcomes underscore the potential of machine learning techniques.
We looked at the delicate balance that exists between these machine learning algorithms' capacities for handling non-image input and image identification.
Our key findings highlight the strengths and distinctions of each algorithm along with their respective accuracies and confusion matrix/graphs.

It is evident from our research that:

- *Random Forest* excels with an accuracy of 96%. The increased accuracy of Random Forest can be attributed to its proficiency in handling both numerical and categorical characteristics, as well as its capacity to capture complicated, non-linear relationships within the data.

- *Convolutional Neural Networks* (CNN) follow up with an accuracy of 93%.CNN performed admirably, but its application to accident prediction may need to be explored further, perhaps by adding more spatial and image-based data.

- *K-Nearest Neighbors* (KNN) demonstrate an accuracy of 91%. Although KNN has an adequate accuracy rate, its predictive potential could be improved by tweaking hyperparameters or feature engineering.

The findings of the study have substantial implications for improving accident prediction systems and may help with proactive actions for accident prevention and preparedness for emergencies preparation.
To increase the robustness of accident prediction models, future research in this field should concentrate on adding real-time data sources, larger datasets, and a wider range of attributes.
In a nutshell, this study adds to the body of knowledge on accident prediction modelling and emphasises the potential of machine learning algorithms, with Random Forest standing out as a promising option for precise accident prediction.
Continued developments in this area have the potential to greatly increase road safety and improve emergency response plans.

Looking ahead, our study on accident prediction using machine learning algorithms, specifically Convolutional Neural Networks (CNN), K-Nearest Neighbors (KNN), and Random Forest (RF), points to several promising directions for future research.

First, there is a need to explore the integration of real-time data sources, such as traffic conditions and weather updates, to enhance the accuracy and responsiveness of prediction models.

Second, expanding datasets to encompass diverse scenarios and accident types could significantly improve the models' generalisation capabilities.

Third, investigating hybrid models that combine the strengths of different algorithms offers potential for increased robustness. Refining machine learning models through feature engineering and hyperparameter tuning represents another avenue for future exploration.

Additionally, the incorporation of advanced image recognition techniques and more diverse visual information in accident prediction models, particularly those utilising CNN, warrants further investigation.
Developing explainable AI techniques will be crucial for understanding and interpreting the decision-making processes of increasingly complex models.
Exploring how accident prediction models can be integrated into autonomous vehicle systems for real-time risk assessment is an emerging area with significant implications for road safety.
Lastly, establishing collaborative efforts with emergency response systems can lead to the development of integrated accident prediction and response systems, enhancing the overall efficiency of emergency services.

In summary, these future research directions hold substantial promise for advancing the sophistication and practicality of accident prediction systems, contributing to the mitigation of the impact of accidents on individuals and society.

References

[1] S. Tan and Y. Guo, "Deep learning for traffic accident prediction," Transportation Research Part C: Emerging Technologies, vol. 89, pp. 205-217, 2018.

[2] C. Johnson and A. Smith, "Comparative analysis of machine learning algorithms for industrial accident prediction," Safety Science, vol. 120, pp. 416-427, 2019.

[3] J. Williams et al., "Human factors in traffic accident prediction: A psychological approach," Accident Analysis & Prevention, vol. 138, p. 105500, 2020.

[4] P. Liu and Q. Wang, "Challenges in real-time accident prediction models," Transportation Research Part C: Emerging Technologies, vol. 126, p. 103225, 2021.

[5]https://www.kaggle.com/datasets/ckay16/accident-detection-from-cctv-footage?resource=download