COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

MASTERS IN MOTORSPORT, MOBILITY AND SAFETY

MASTERS THESIS

DEVELOPMENT OF A DATABASE FOR CRASH SEVERITY EVALUATION BASED ON REAL CRASH DATA

Author: José María de Inza García

Director: Malte Plenter

Co-Director: Michael Duering

Wolfsburg

August 2025

Content

Chapter 1. INTRODUCTION

Road traffic crashes are a major global public health challenge, causing approximately 1.19 million deaths annually, along with millions of severe injuries, long-term disabilities, and substantial economic costs (World Health Organization, 2024). Accurate assessment of crash severity is therefore critical for improving vehicle safety, guiding regulations, and advancing crash reconstruction science.

This thesis addresses this challenge by refining crash severity evaluations through improved estimation of Energy Equivalent Speed (EES), a key parameter that quantifies the energy dissipated during a collision. Traditional EES estimation methods are often limited by simplifications or incomplete data. To overcome these limitations, this work leverages detailed real-world data from the German In-Depth Accident Study (GIDAS), enabling the development of methodologies that yield EES values more accurately reflecting collision dynamics.

A central focus of the research is partner protection, the principle that vehicles should protect not only their occupants but also minimize injury risk to occupants of other vehicles. By using precise EES codings, this thesis highlights disparities in energy dissipation among different vehicle types and sizes, emphasizing the importance of design strategies that balance protection across all traffic participants.

The outcomes align with international safety initiatives, including Euro NCAP and the U.S. NHTSA, supporting more reliable safety assessments and crash test protocols. In addition to improving vehicle design, the refined methodologies have broader applications in forensic crash reconstruction, policy-making, and the harmonization of global safety standards.

INTERNAL

Chapter 2. DEFINITION OF THE PROJECT

This project aims to improve the accuracy of crash severity assessments by developing a structured database that recalibrates Energy Equivalent Speed (EES) values using real-world crash data from the German In-Depth Accident Study (GIDAS) (Otte, 2003). EES quantifies the kinetic energy absorbed during a collision, serving as a crucial indicator of crash severity. Existing methods for determining EES, however, often rely on subjective coding, which can introduce inconsistencies, particularly when based on limited or non-standardized visual evidence.

To address these challenges, the project introduces a tool capable of estimating EES retrospectively by combining scene photographs with vehicle-specific parameters such as geometry, structural characteristics, and deformation patterns. This data-driven approach reduces reliance on subjective visual assessments and produces more consistent and reproducible EES values across different crash scenarios. By grounding estimations in objective information, the tool enhances the reliability of crash severity evaluations and supports broader applications in accident reconstruction and vehicle safety research.

Another goal of the project is to identify factors contributing to coding errors, including inconsistencies in visual interpretation, variations in crash configuration, and differences in vehicle mass and structure. By quantifying these sources of error, the project strengthens crash analysis methodologies and provides clearer guidance for improving data quality. The refined EES values also allow deeper examination of partner protection—ensuring vehicle design safeguards all collision participants—and the mass ratio effect, which can increase injury risk for smaller vehicles in collisions with larger ones.

Overall, this initiative advances both scientific understanding and practical traffic safety outcomes. It enables more reliable crash reconstructions, informs safer vehicle design, and supports evidence-based regulatory decisions. By aligning with international safety programs such as Euro NCAP and NHTSA, the project contributes to global efforts to improve road safety and harmonize vehicle safety standards.

INTERNAL

Chapter 3. DESCRIPTION OF THE TOOLS

3.1 EES ANALYZER

As part of this research, a new tool called EES Analyzer was developed with the primary purpose of providing a comprehensive framework for analyzing the GIDAS database. The tool enables systematic comparison of accident cases, the identification of crashes with similar configurations, and the exploration of relationships among variables associated with crash severity. One of its most significant features is its ability to predict EES values based on selected parameters, which makes it possible to detect inconsistencies in coding and thereby improve the overall reliability of the dataset.

The graphical user interface was implemented in R Shiny, ensuring an interactive and adaptable environment for researchers. This interface allows cases to be filtered by crash configuration, vehicle model, category, and EES ranges, providing a high degree of flexibility in case selection. In addition, VDI filters (VDI1, VDI2, and VDI3) were incorporated to classify accidents lacking Brumbelow coding, thus expanding the number of cases that can be included in the analysis. The system also offers options to exclude motorcycles, trucks, and multiple collisions, with a focus on simple frontal car-to-car crashes, which ensures greater consistency and comparability across cases.

INTERNAL

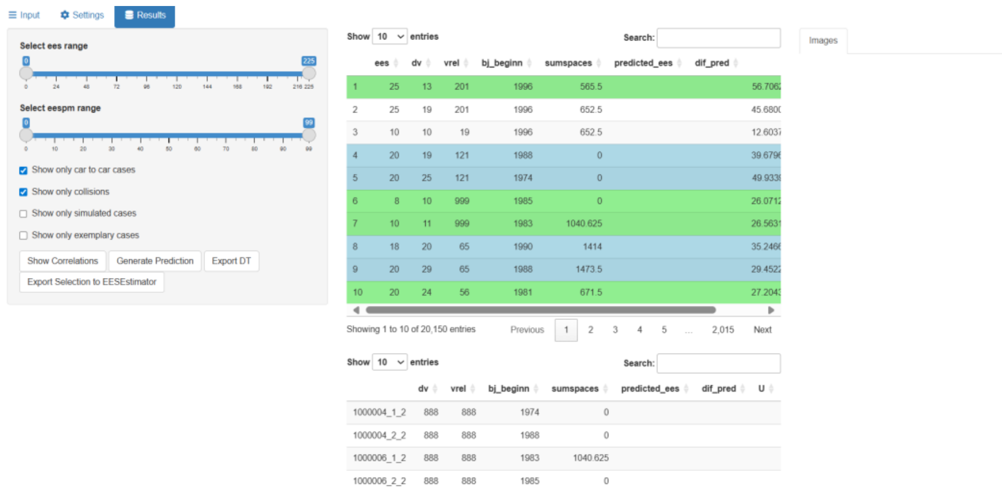


Figure 1: Output Table of EES Analyzer

The EES Analyzer integrates both exploratory and predictive functions. Linear regressions are used to investigate correlations between EES and other crash variables, while machine learning models, specifically Random Forest, generate predicted EES values that can be contrasted with the coded ones in GIDAS. This approach allows the identification of potential outliers and possible miscoded cases. Results are displayed through a set of diagnostic visualizations, including correlation matrices, scatter plots, heatmaps, and box plots of prediction errors, which facilitate both interpretation and validation of findings.

In the process of developing the tool, several new variables were derived to enrich the analysis. These included the total deformation area, calculated from zones F1–F4, although it showed only limited correlation with EES ($R^2 = 0.12$). A second derived parameter was the post-impact velocity, computed from Delta-V, impulse angle, and angular change, which served as an intermediate variable for energy-based calculations. The most relevant contribution, however, was the definition of an equivalent velocity (U), obtained from dissipated kinetic energy and the mass ratio between vehicles. This new variable demonstrated a strong correlation with EES ($R^2 = 0.76$), becoming the most consistent and robust predictor across configurations.

Based on these inputs, a Random Forest model was trained to estimate EES values more accurately. The model incorporated relative velocity, vehicle year, total deformation, and equivalent velocity (U) as the main predictors. Among them, U consistently emerged as the most

INTERNAL

influential variable, while vehicle year showed comparatively lower relevance. The model was trained on all available cases with valid EES values, allowing not only prediction but also the detection of outliers that could be iteratively corrected to improve database coherence. Performance was evaluated using confusion matrices, cross-validation with ten folds, and graphical error analysis, confirming the robustness of the proposed approach.

Ultimately, the EES Analyzer aims to detect and correct errors in EES coding for frontal car-to-car crashes, thereby generating cleaner and more reliable datasets. Beyond its immediate application, the tool holds the potential to be scaled to the analysis of all crash types in the GIDAS database. This scalability underscores its relevance for both research and applied safety studies, as it provides a structured, automated, and scientifically grounded framework for improving crash severity assessments.

3.2 EES ESTIMATOR

Once potential miscoded cases are identified, the EES Estimator tool is used to generate an estimated EES value for each case. These estimated values are subsequently compared with both the coded EES values in the GIDAS database and the predicted values obtained from the EES Analyzer tool. This comparative analysis allows researchers to evaluate the accuracy of coding, detect inconsistencies, and draw informed conclusions about the reliability of the data.

The EES Estimator was developed as part of this thesis and integrates the EES-CNN functionality of PC-Crash with Microsoft Copilot for automated image recognition. The primary goal of this tool is to automate the repetitive process of uploading case images to PC-Crash, retrieving the EES estimations, and saving the results without requiring continuous user intervention. Conceptually, the tool operates by controlling the mouse and keyboard through R, executing a predefined sequence of actions that adapts dynamically based on program inputs.

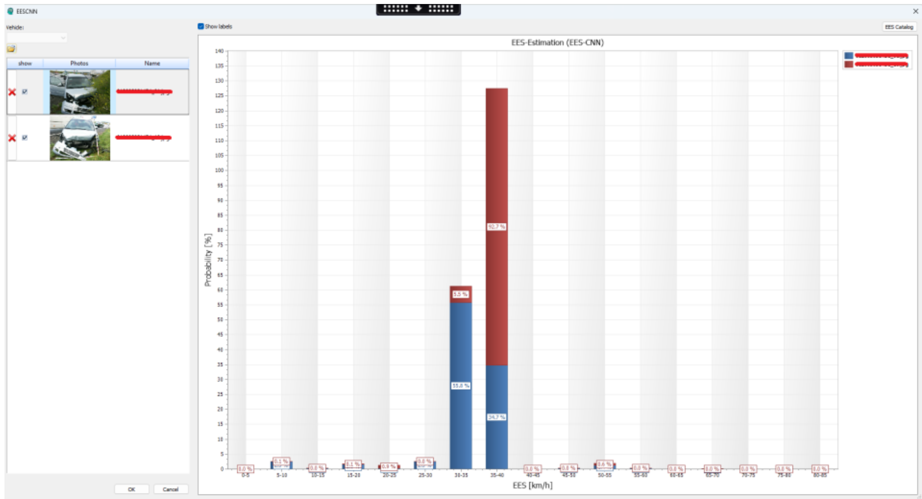


Figure 2: PC-Crash Visual Estimation Output

The process begins with the preparation of the environment, where Microsoft Copilot is opened while all other programs, except RStudio, are closed. The program then opens Copilot and retrieves two selected images for each crash case, identified as the most relevant for damage assessment. Copilot validates the images automatically to ensure they can be used in the analysis. Following this, PC-Crash is launched, the EES-CNN tool is accessed, and the first image is loaded into the system. Upon completing the EES estimation, PC-Crash generates a graph indicating the likely EES range along with a confidence level. Because PC-Crash does not export numerical outputs, the tool automatically takes a screenshot of the graph for further processing.

The screenshots are then analyzed using Microsoft Copilot, which extracts the data from the image. The program repeats this process for the second image, representing a different perspective of the most damaged area. Copilot compares the results from both images, calculates the average for each bar in the graphs, and returns the estimated EES range and confidence level in a structured format. These values are subsequently stored in the input dataset as variables labeled “min_eescnn,” “max_eescnn,” and “conf_eescnn.”

The validation of uploaded images is critical, as mislabeling or misplacement of photos can occur. Some images may depict irrelevant scenes, incorrect angles, or even missing content, which could reduce the confidence of the resulting estimates. The EES Estimator addresses these issues through automated checks and the use of multiple images per case to improve reliability.

This process can be repeated for any number of cases specified by the user. After completing the estimation for each case, the program updates an Excel table, allowing for direct comparison of coded EES values in GIDAS, predicted values from the EES Analyzer, and calculated values from the EES Estimator. While the EES Estimator operates as a separate program from the EES Analyzer, the Analyzer includes an export feature that allows up to 1,000 rows of data to be processed, prioritizing cases with the largest discrepancies between coded and predicted EES. This limit was chosen to balance computational time and study objectives, as each estimation takes approximately two minutes, making the processing of 1,000 cases feasible within roughly 34 hours.

Chapter 4. RESULTS

This study evaluated the extent to which crash configuration influences the accuracy of EES prediction by applying a Random Forest model to a dataset of 2,288 frontal crashes, which represent a subset of the 20,000 cases included in the global study. By disaggregating the analysis according to crash configuration, the study was able to uncover important differences in both variable importance and model performance. A newly derived parameter, equivalent velocity—obtained from kinetic energy calculations—emerged as the most consistent and robust predictor across all configurations, underscoring its value as a core input for EES modeling.

The results revealed that model accuracy and stability varied significantly depending on the crash configuration. For instance, oblique center impacts tended to present a higher proportion of potential outliers and larger prediction errors, suggesting greater complexity in capturing their dynamics. Similarly, small and moderate overlap crashes exhibited greater dispersion at higher EES values, indicating that the prediction model struggles to capture non-linear behaviors in these crash types. These findings demonstrate that a "one-size-fits-all" approach to EES prediction may overlook critical nuances in crash geometry and energy distribution.

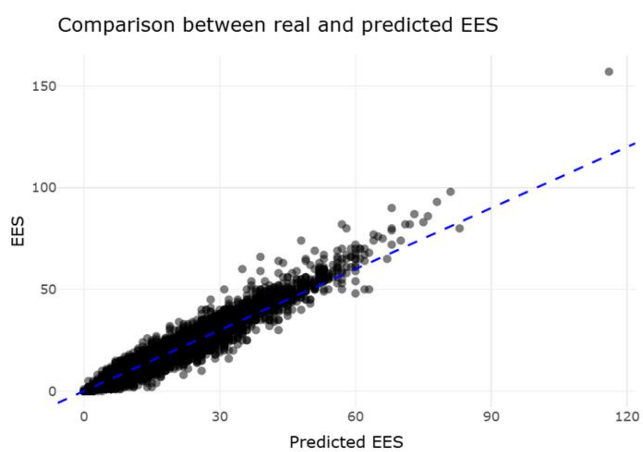


Figure 3: Comparison between GIDAS coded EES and EES Analyzer Predicted EES

The comparison highlights divergences between coded values and model predictions, with noticeable clusters of miscoded or misestimated cases.

To strengthen the reliability of EES predictions, a multi-stage correction process was implemented. This included advanced filtering techniques, systematic outlier detection, and validation using PC-Crash simulations. By triangulating results between GIDAS-coded EES values, EES Analyzer predictions, and PC-Crash estimations, the study generated a clearer picture of the alignment and deviations among different methods. The integration of scatter plots and heatmaps provided a detailed visual representation of these relationships, offering insights into both systematic biases and random errors.

The overall findings emphasize the critical role of crash configuration in refining EES assessment methodologies. By identifying where prediction models perform well and where they encounter limitations, the study provides a foundation for developing more configuration-sensitive models. Such improvements not only enhance the scientific rigor of crash reconstruction but also contribute to vehicle safety analysis, particularly in areas such as partner protection design, occupant injury mitigation, and the calibration of safety testing protocols.

Figure 4: Comparison of EES between GIDAS, EES Analyzer Prediction and EES Estimator

Chapter 5. CONCLUSIONS

This study investigated inaccuracies in EES coding within the GIDAS database, focusing on the interplay between crash characteristics, vehicle type, and prediction performance. The analysis revealed that crash configuration had only a limited influence on prediction errors, suggesting that other factors play a more dominant role in misestimations. In particular, higher EES values and larger vehicle categories, such as SUVs, were more prone to inaccurate codings. These results point to potential biases in the coding process, where extreme values and structurally different vehicle types introduce greater uncertainty.

A notable finding was the strong correlation between EES and Delta-V, which raises important concerns regarding the independence of the two metrics. Since Delta-V is already a widely used measure of crash severity, excessive reliance on it in EES codings could undermine the distinctiveness and added value of EES as an independent indicator. This observation highlights the need for more rigorous methodologies to ensure that EES retains its intended role as a complementary, physics-based measure of crash severity.

When comparing modeling approaches, generic prediction models occasionally outperformed configuration-specific ones, indicating that broader models may be more resilient to data variability and less sensitive to inconsistencies in crash categorization. At the same time, the use of visual diagnostic tools—such as scatter plots and heatmaps—proved effective in detecting potential outliers by contrasting observed values against model predictions.

The combined application of the Random Forest model with the EES Estimator demonstrated strong potential for identifying and correcting errors within the database. This hybrid approach provides a systematic way of improving the reliability of EES codings by integrating machine learning prediction, reference estimation, and validation procedures.

INTERNAL

Chapter 6. RELATION WITH M2S

The coursework undertaken in this thesis provided a solid foundation for understanding vehicle safety, crash reconstruction, and mobility systems. The Crashworthiness course introduced global initiatives such as Euro NCAP, emphasizing the relevance of Energy Equivalent Speed (EES) in vehicle design and occupant protection, and offered conceptual grounding through MADYMO simulations for interpreting vehicle kinematics.

Data Analysis and Visualization with Python equipped the research with skills for managing large datasets, structuring crash data, and creating clear visualizations, directly supporting the EES Analyzer database and the interpretation of coded, predicted, and simulated EES values. Global Transportation offered insight into political and regulatory frameworks, facilitating the use of authoritative crash data sources such as GIDAS and situating the thesis within international road safety objectives. Integrated Safety and Restraint Systems strengthened understanding of vehicle kinematics and energy dissipation, linking crash severity analysis to occupant protection.

The Telemetry and Data Acquisition course provided a framework to understand the reliability and limitations of vehicle data, indirectly supporting critical engagement with EES coding. Injury Biomechanics explained how Delta-V and EES relate to human injury tolerance, highlighting the importance of accurate estimations for predicting occupant outcomes. Sustainable Mobility introduced systemic approaches to safety, aligning with partner protection by considering all road users, while Vehicle Dynamics offered technical grounding to interpret crash kinematics and informed statistical model design. Composites and Lightweight provided context on vehicle structural deformation and energy absorption, enriching the interpretation of real-world crash outcomes.

Collectively, these courses bridged theoretical knowledge, technical methodology, and societal perspectives, equipping the thesis with the tools necessary to advance crash severity assessment and contribute meaningfully to vehicle safety research.

Chapter 7. BIBLIOGRAPHY

European New Car Assessment Programe (Euro NCAP). (2017). EuroNCAP 2025 Roadmap. Von https://www.euroncap.com/en/for-engineers/technical-papers/ abgerufen

Otte, D. (2003). *Scientific Approach and Methodology of a New In-Depth-Investigation Study in Germany so called GIDAS.*

World Health Organization. (2024). *Global Status Report on Road Safety 2023.*

ANEX I: COMPLETE MII

THESIS

COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MASTERS IN INDUSTRIAL ENGINEERING

MASTERS THESIS

DEVELOPMENT OF A DATABASE FOR CRASH SEVERITY EVALUATION BASED ON REAL CRASH DATA

Author: José María de Inza García

Director: Malte Plenter

Co-Director: Michael Duering

Madrid

August 2025

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título.

Development of a Database for Crash Severity Evaluation Based on Real Crash Data

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2024/25 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

de Inza Jose Maria VWPKI E167ECCF47D 0359D

Digitally signed by de Inza Jose Maria VWPKI E167ECCF47D0359D Date: 2025.08.26 14:43:23 +02'00'

Fdo.: José María de Inza Fecha: 26 / 08 / 2025

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Plenter Malte VWPKI 4CDD667AD8A 934E9

Digital unterschrieben von Plenter Malte VWPKI 4CDD667AD8A934E9 Datum: 2025.08.26 15:04:02 +02'00'

Fdo.: Malte Plenter Fecha: 26 / 08 / 2025

Disclaimer

The results, opinions and conclusions of this thesis are not necessarily those of Volkswagen AG.

MASTERS IN INDUSTRIAL ENGINEERING

MASTERS THESIS

DEVELOPMENT OF A DATABASE FOR CRASH SEVERITY EVALUATION BASED ON REAL CRASH DATA

Author: José María de Inza García

Director: Malte Plenter

Co-Director: Michael Duering

Madrid

August 2025

Acknowledgements

Primarily, I would like to express my sincere gratitude to my thesis supervisor, Malte Plenter, for his continuous guidance, insightful feedback, and support throughout this project. I am equally grateful to Dr. Michael Duering, who, as my co-supervisor and internship supervisor, provided valuable opportunities and guidance during my practical work experience.

I would also like to extend my thanks to the University, who gave me everything I ever needed to learn and grow, and to the Chair of Motorsport, Mobility, and Safety, whose organization of the second Master's program and assistance in securing my internship placement made this thesis possible.

This work was partially supported by an Erasmus+ scholarship, which greatly helped me during my relocation to Wolfsburg.

I am deeply thankful to the entire Crash Research Team at Volkswagen, whose expertise and readiness to assist were invaluable during the development of this thesis. Their support, whether in technical IT matters or in strategic discussions regarding the study's progression, was of immense help. I would also like to acknowledge my fellow interns from the Crash Research Team, with whom I had the pleasure of sharing an office and daily experiences throughout this journey.

On a personal note, I owe my deepest gratitude to my family — my father, mother, grandmother, aunt, sister, and all the rest of my family — for their unconditional support, encouragement, and patience during the entire process. Their belief in me has been my greatest source of strength. I also want to thank my group of friends from the last years of university, whose companionship and support made this academic journey both enriching and unforgettable.

This thesis marks the culmination of a journey filled with learning, challenges, and growth, which would not have been possible without the invaluable support of all those mentioned above.

DESARROLLO DE UNA BASE DE DATOS PARA EL ESTUDIO DE LA GRAVEDAD DE COLISIONES BASADA EN DATOS DE COLISIONES REALES.

Autor: de Inza García, José María.
Director: Plenter, Malte.
Entidad Colaboradora: Volkswagen

RESUMEN DEL PROYECTO

Esta investigación desarrolla una herramienta basada en datos para mejorar la precisión en la evaluación de la gravedad de las colisiones, centrándose en la refinación de la Velocidad Equivalente de Energía (EES) utilizando datos reales. Destaca las limitaciones de las prácticas actuales de codificación, el impacto de las configuraciones de choque y el valor de los modelos predictivos para identificar y corregir casos mal codificados.

Keywords: EES, GIDAS, Colisiones de Trafico, Protección del Compañero de Choque, Base de datos.

1. Introducción

Las colisiones de tráfico causan aproximadamente 1,19 millones de muertes anualmente en todo el mundo. (World Health Organization, 2024). Esta tesis aborda el problema mediante la mejora de las evaluaciones de la gravedad del choque, lo cual permite el desarrollo de sistemas de seguridad mas preciso, refinando la estimación de la Velocidad Equivalente de Energía (EES) con datos reales extraídos de la base GIDAS. También promueve el concepto de protección del vehículo contrario en el diseño automotriz, resaltando disparidades en la disipacion de energia mediante valores de EES precisos, y se alinea con iniciativas globales de seguridad como Euro NCAP y NHTSA, contribuyendo a una reconstrucción de choques más fiable y a una movilidad más segura. (European New Car Assessment Programe (Euro NCAP), 2017).

2. Definition of the Project

Este proyecto busca mejorar la precisión en la evaluación de la gravedad de los choques mediante el desarrollo de una base de datos estructurada y completa que recalibra los valores de Velocidad Equivalente de Energía (EES) utilizando datos reales de colisiones extraídos de GIDAS (Otte, 2003). Introduce una herramienta capaz de estimar retrospectivamente la Velocidad Equivalente de Energía (EES) a partir de fotografías de la escena y parámetros específicos del vehículo, abordando las limitaciones de las prácticas actuales de codificación que dependen en gran medida de evaluaciones visuales subjetivas. La iniciativa también busca identificar los factores que contribuyen a errores de codificación y mejorar el análisis de la protección del vehículo contrario y los efectos de la relación de masas, apoyando reconstrucciones de choques más fiables y decisiones más seguras en diseño vehicular y regulación.

3. Descripción de las Herramientas

La herramienta EES Analyzer permite una comparación y análisis eficiente de los casos de choque en la base de datos GIDAS, incluyendo la predicción de valores EES basada en variables seleccionadas para ayudar a detectar inconsistencias.

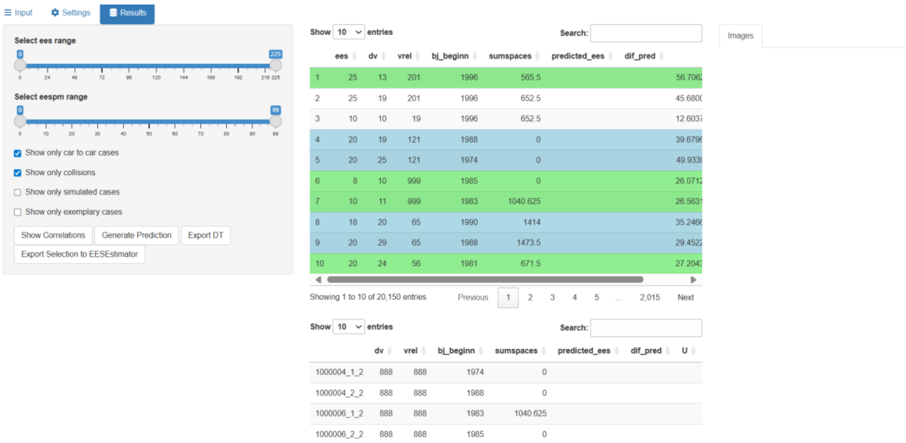


Ilustración 1: Tabla Generada por EES Analyzer

Una vez identificados los posibles casos mal codificados, la herramienta EES Estimator genera un valor de referencia de EES, que luego se compara con el valor codificado en GIDAS y el valor predicho por EES Analyzer. Este proceso se automatiza mediante la integración de GIDAS, Microsoft Copilot y PC-Crash.

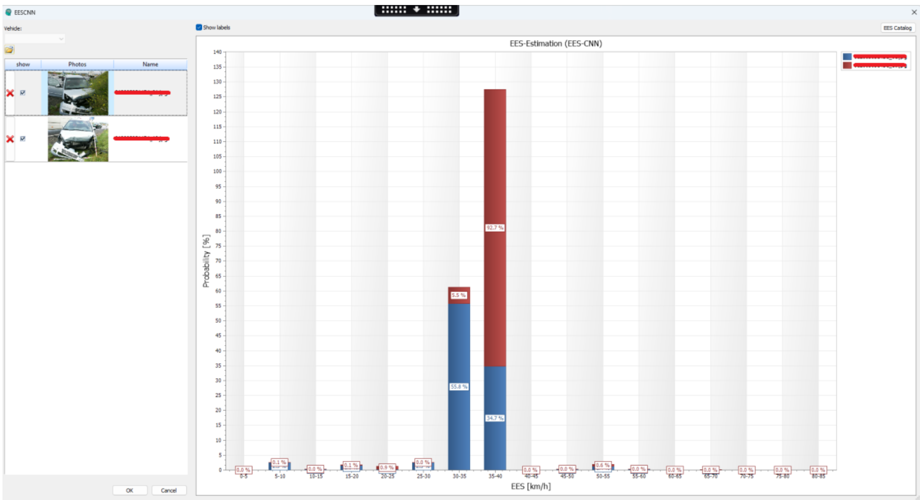


Ilustración 2: Estimación de EES mediante inspección visual con PC-Crash

4. Resultados

Este estudio evaluó cómo la configuración del choque influye en la precisión de la predicción de EES utilizando un modelo Random Forest aplicado a 2.288 colisiones frontales de las 20.000 que se han empleado para el studio global. Al analizar cada configuración por separado, se observaron diferencias en la importancia de las variables y el comportamiento del modelo, siendo la velocidad equivalente el predictor más constante. Algunas configuraciones, como el choque oblicuo central, mostraron más valores atípicos y errores, mientras que los solapamientos pequeños y moderados presentaron mayor dispersión en valores altos de EES.

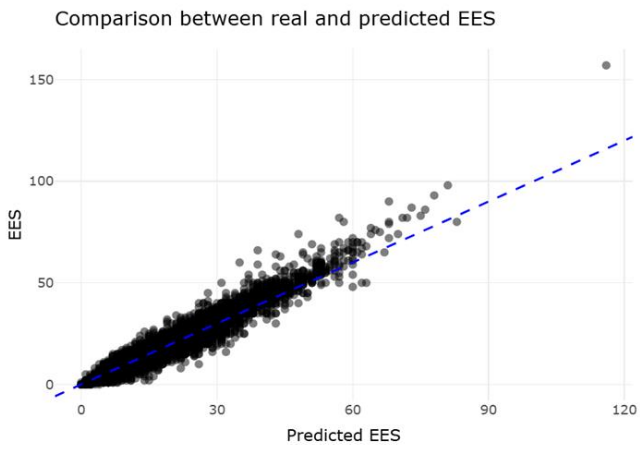


Ilustración 3: Comparativa entre el EES definido en GIDAS y el predicho por EES Analyzer

Para mejorar la fiabilidad, se aplicó un proceso de corrección que incluyó filtrado, detección de valores atípicos y validación con PC-Crash. La comparación entre el EES codificado en GIDAS, las predicciones de EES Analyzer y las estimaciones de PC-Crash mostró distintos niveles de coincidencia, visualizados mediante gráficos de dispersión y mapas de calor. Los resultados destacan el papel de la configuración del choque en la mejora de las evaluaciones de EES y el análisis de seguridad vehicular.

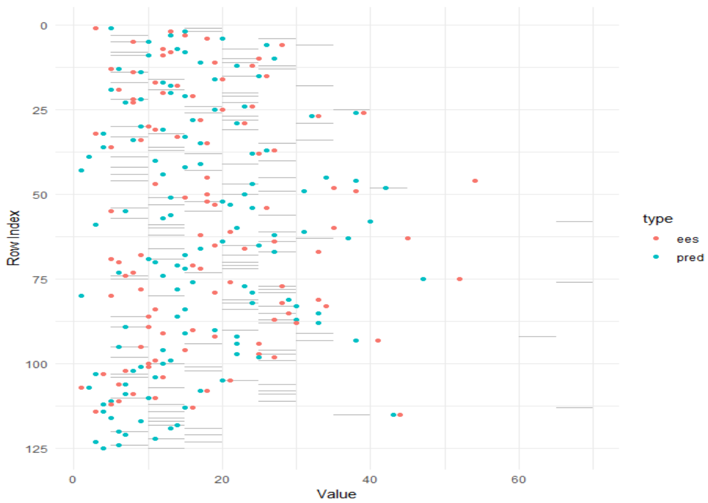


Ilustración 4: Comparativa entre el EES definido en GIDAS, el predicho por EES Analyzer y el estimado con EES Estimator

5. Conclusiones

Este estudio analizó las inexactitudes en la codificación de EES dentro de la base de datos GIDAS. La configuración del choque mostró un impacto limitado en los errores de predicción, mientras que los valores altos de EES y los vehículos grandes como los SUV fueron más propensos a errores. Una fuerte correlación entre EES y Delta-V plantea dudas sobre la independencia de estas métricas, ya que la codificación de EES podría estar dependiendo más de esta variable de lo adecuado.

Los modelos genéricos superaron en algunos casos a los específicos por configuración, y las herramientas visuales ayudaron a detectar valores atípicos. El modelo Random

Forest combinado con EES Estimator resultó útil para corregir errores. El trabajo futuro se centrará en depurar los datos y automatizar las predicciones de EES dentro de EES Analyzer.

6. Referencias

European New Car Assessment Programe (Euro NCAP). (2017). EuroNCAP 2025 Roadmap. Von https://www.euroncap.com/en/for-engineers/technical-papers/ abgerufen

Otte, D. (2003). *Scientific Approach and Methodology of a New In-Depth-Investigation Study in Germany so called GIDAS.*

World Health Organization. (2024). *Global Status Report on Road Safety 2023.*

DEVELOPMENT OF A DATABASE FOR CRASH SEVERITY EVALUATION BASED ON REAL CRASH DATA.

Author: de Inza García, José María.
Supervisor: Plenter, Malte.
Collaborating Entity: Volkswagen

ABSTRACT

This research develops a data-driven tool to improve the accuracy of crash severity assessments, focusing on refining Energy Equivalent Speed using real-world data. It highlights the limitations of current coding practices, the impact of crash configurations, and the value of predictive modeling for identifying and correcting miscoded cases.

Keywords: EES, GIDAS, Traffic crash, Partner protection, Data base.

1. Introduction

Road traffic crashes cause approximately 1.19 million deaths annually worldwide (World Health Organization, 2024). This thesis addresses the issue by improving crash severity assessments, leading to more accurate safety systems development, through a refined estimation of Energy Equivalent Speed (EES), using real-world data from the GIDAS database. It also promotes the concept of partner protection in vehicle design, highlighting energy dissipation disparities through accurate EES codings, and aligns with global safety initiatives like Euro NCAP and NHTSA, contributing to safer roads and more reliable crash reconstruction practices (European New Car Assessment Programe (Euro NCAP), 2017).

2. Definition of the Project

This project seeks to enhance the precision of crash severity evaluations by developing a comprehensive, structured database that recalibrates Energy Equivalent Speed (EES) values using real-world crash data from GIDAS (Otte, 2003). It introduces a tool capable of estimating EES retrospectively from scene photographs and vehicle-specific parameters, addressing the limitations of current coding practices that rely heavily on subjective visual assessments. The initiative also aims to identify factors contributing to coding errors and improve the analysis of partner protection and mass ratio effects, ultimately supporting more reliable crash reconstructions and informing safer vehicle design and regulatory decisions.

3. Description of the Tool

The EES Analyzer tool allows efficient comparison and analysis of crash cases in the GIDAS database, including the prediction of EES values based on selected variables to help detect inconsistencies.

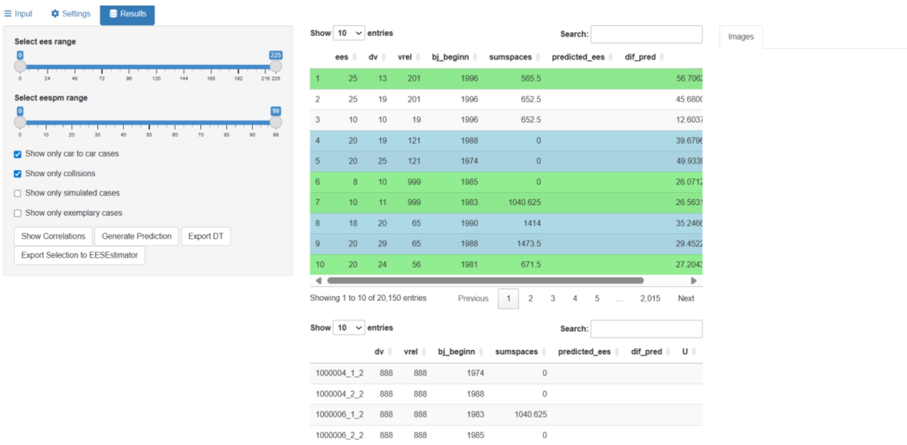


Figure 1: Output Table of EES Analyzer

Once potential miscoded cases are found, the EES Estimator generates a reference EES value, which is then compared with both the GIDAS-coded and predicted values. This process is automated through the integration of GIDAS, Microsoft Copilot, and PC-Crash.

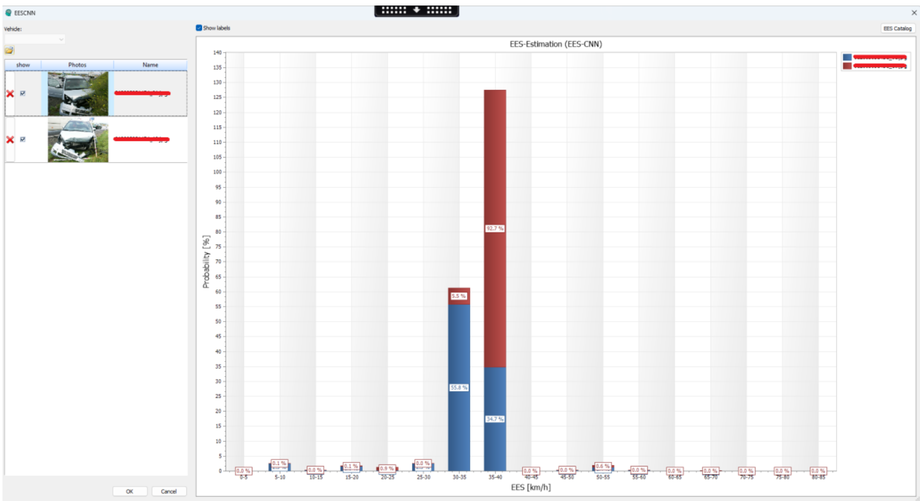


Figure 2: PC-Crash Visual Estimation Output

4. Results

This study evaluated how crash configuration influences EES prediction accuracy using a Random Forest model applied to 2,288 frontal crashes, which belong to the 20.000 cases of the global study. By analyzing each configuration separately, it revealed differences in variable importance and model behavior, with a new variable obtained from kinetic energy, defined as equivalent velocity, being the most consistent predictor. Certain configurations, like oblique center, showed more potential outliers and errors, while small and moderate overlaps had greater dispersion at higher EES values.

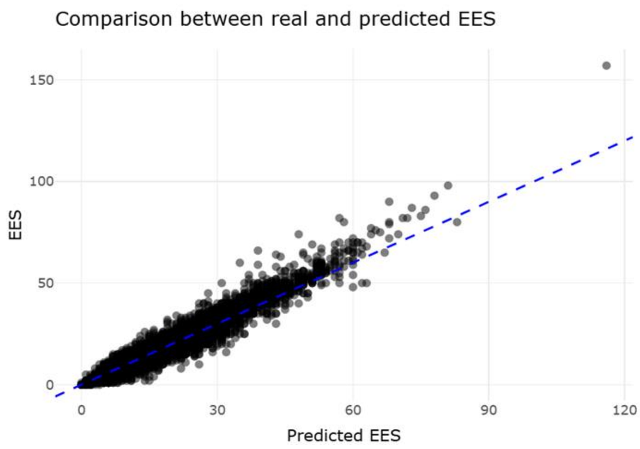


Figure 3: Comparison between GIDAS coded EES and EES Analyzer Predicted EES

To improve reliability, a correction process was applied using filtering, outlier detection, and PC-Crash validation. A comparison between GIDAS-coded EES, EES Analyzer predictions, and PC-Crash estimations showed varying degrees of alignment, visualized through scatter plots and heatmaps. The findings emphasize the role of crash configuration in refining EES assessments and improving vehicle safety analysis.

Figure 4: Comparison of EES between GIDAS, EES Analyzer Prediction and EES Estimator

5. Conclusions

This study analyzed inaccuracies in EES coding within the GIDAS database. Crash configuration showed limited impact on prediction errors, while higher EES values and larger vehicles like SUVs were more prone to misestimation. A strong correlation between EES and Delta-V raised concerns about metric independence, meaning that EES codings could be relying more on that variable than they should.

Generic models sometimes outperformed configuration-specific ones, and visual tools helped detect potential outliers against the prediction of the model. The Random Forest model combined with EES Estimator proved useful for correcting errors. Future work will focus on cleaning the data and automating EES predictions within the EES Analyzer.

References

European New Car Assessment Programe (Euro NCAP). (2017). EuroNCAP 2025 Roadmap. Von https://www.euroncap.com/en/for-engineers/technical-papers/ abgerufen

Otte, D. (2003). *Scientific Approach and Methodology of a New In-Depth-Investigation Study in Germany so called GIDAS.*

World Health Organization. (2024). *Global Status Report on Road Safety 2023.*

Content of the Thesis

Index of Figures

Index of Tables

Chapter 1. INTRODUCTION

Road traffic safety is a critical issue worldwide, with traffic crashes consistently ranking among the leading causes of death and severe injury. According to the World Health Organization, approximately 1.3 million people die each year as a result of road traffic crashes, and millions more suffer non-fatal injuries (World Health Organization, 2024). These figures highlight the urgent need for improved vehicle safety measures and more accurate tools for analyzing crash severity.

This thesis addresses this challenge by developing a comprehensive crash severity database based on real-world accident data. The primary goal is to improve the accuracy of crash severity assessments, with a particular focus on the estimation of Energy Equivalent Speed (EES).

EES is a derived metric used in accident reconstruction to estimate the severity of a vehicle's structural deformation in a crash. In the context of real-world collisions, EES represents the speed at which a vehicle would need to impact an object to dissipate an equivalent amount of energy through deformation as observed in the actual crash. Unlike in controlled crash tests, where EES is often calculated using empirical stiffness coefficients, real-world EES estimation typically relies on alternative methodologies, such as expert judgment, pattern recognition, or data-driven reconstruction techniques, due to the variability and complexity of crash scenarios.

While EES is not a direct measure of pre-impact speed, it is influenced by it, particularly in cases where the energy absorbed by the vehicle structure correlates with the kinetic energy involved in the collision. Therefore, EES serves as a proxy for crash severity, reflecting the energy dissipated through vehicle deformation, but it should not be interpreted as a substitute for actual impact velocity.

In addition to refining EES calculations, this project explores the concept of partner protection, which is the idea that vehicles should be designed not only to protect their own occupants but also to minimize the damage they inflict on other vehicles in a collision. This concept is particularly relevant in car-to-car crashes involving vehicles of different sizes or masses. Although not yet widely standardized, partner protection is increasingly recognized as a key aspect of vehicle compatibility and overall road safety. Recent initiatives, such as the U.S. Department of Transportation's 2025 Progress Report on the National Roadway Safety Strategy, highlight the growing focus on protecting all road users through safer vehicle design and collaborative safety efforts (United States Department of Transportation, 2025)

Over the past decades, various methodologies have been developed to assess crash severity, incorporating factors such as vehicle deformation, impact speed, mass, and structural compatibility. However, many of these approaches lack consistency and fail to account for the complexity of real-world collisions. This project builds on previous work by creating a detailed and validated crash severity database, re-evaluating EES values using reference cases, and identifying sources of error in current EES coding practices.

The analysis is based on data from the German In-Depth Accident Study (GIDAS), which provides high-resolution information on car-to-car crashes. These data enable the investigation of crash scenarios involving different mass ratios and structural behaviors. By improving the quality of EES coding, the project aims to enhance our understanding of crash severity and partner protection, ultimately contributing to safer vehicle designs and more effective safety regulations.

By achieving these objectives, this project aims to contribute meaningfully to the reduction of traffic-related fatalities and injuries. The insights gained from analyzing real-world crash data can support the development of more robust vehicle compatibility assessments and inform regulatory bodies and manufacturers about structural mismatches that may compromise safety. These findings align with broader international efforts to enhance vehicle safety standards, such as the European New Car Assessment Programme (Euro

NCAP) and the U.S. National Highway Traffic Safety Administration (NHTSA), which increasingly emphasize not only occupant protection but also inter-vehicle crash compatibility (NHTSA Announces Model Year 2025 Vehicles for 5-Star Safety Ratings Testing, 2024). Ultimately, this research supports the long-term vision of a safer road transport system by providing evidence-based tools and methodologies that can guide both policy and design improvements.

1.1 MOTIVATION OF THE PROJECT

The motivation for this project stems from the critical need to enhance road traffic safety through analyses of field crashes, which profit from accurate assessment of the medical and technical crash severity. Traffic crashes remain one of the leading cause of fatalities and injuries worldwide (World Health Organization, 2024), making it imperative to develop more precise methods for evaluating crash severity. The concept of Energy Equivalent Speed (EES) has been central to this effort, providing a standardized measure of the energy dissipated through deformation by vehicles during collisions.

Despite all advancements achieved with previous studies, challenges remain in achieving a universally reliable method for EES calculation. Existing methods often face issues with noise and irregularities in the data—such as incomplete deformation measurements, inconsistent vehicle documentation, or variability in crash reconstruction inputs—which can introduce significant uncertainty into the estimation process. These factors can distort the energy absorption profile of the vehicle, leading to inaccurate or non-reproducible EES values, particularly in complex or borderline cases.

The ultimate goal of this project is to contribute to the reduction of traffic fatalities and enhance overall road safety. By improving the technical assessment of crash severity, this project aligns with the broader objective of developing safer vehicles and more effective safety regulations, benefiting all road users.

Chapter 2. DESCRIPTION OF TECHNOLOGIES

The successful implementation of this project requires a variety of tools and resources to ensure accurate data collection, analysis, and application. Below is an overview of the necessary resources, which will support the techniques and procedures outlined in the methodology.

2.1 R PROGRAMMING LANGUAGE IN RSTUDIO

R is a powerful tool for statistical computing and data analysis. It was used to extract and prepare data from the GIDAS database. Using the Shiny library, an intuitive HTML-based GUI has been developed to facilitate interaction with the data. While the same functionality could have been implemented without Shiny, the library significantly enhances usability and user experience. R will also be used for statistical analysis of the data, for example, to validate different hypotheses or to check correlations. The most relevant packages will be explained further in the following sections:

2.1.1 SHINY

Shiny is an R package that offers an easy option to create an interactive web application. This package has a simple structure, which is based on three different components:

- The UI: First, the layout of the app is defined. Here, the programmer has to define the visible aspects of the application, defining the tools that the program must offer.
- The server: In a Shiny application, once the user interface (UI) has been defined, the underlying functionality of the tool (including how inputs are processed, how outputs are generated, and how the application reacts to user interactions), needs to be developed in the server component. This part of the application contains the logic that connects the UI elements with the data and computations behind them. Without this server-side programming, the UI remains static and non-functional.

- Call ShinyApp Function: After defining both the server and the UI components, the Shiny application is launched by calling the shinyApp() function. This basic approach is sufficient for running simple applications. However, in more advanced scenarios, the app can be structured and launched using custom packages or modularized code, allowing for better scalability, maintainability, and integration into larger R projects.

This package will allow the development of the EES Analyzer tool in a visually appealing and user-friendly way.

2.1.2 RANDOMFOREST

The RandomForest package is a machine learning algorithm, which allows the prediction of variables by generating decision trees. The algorithm generates different decision trees with random parts of the total data. Each tree in the random forest is unique, as it is trained on a different random subset of the data (rows) and considers a random subset of features (columns) at each split. This randomness ensures diversity among the trees, which contributes to the overall robustness and accuracy of the ensemble model. After generating all the trees and developing a prediction with each separate tree, the algorithm will combine the predictions generated to get a final prediction.

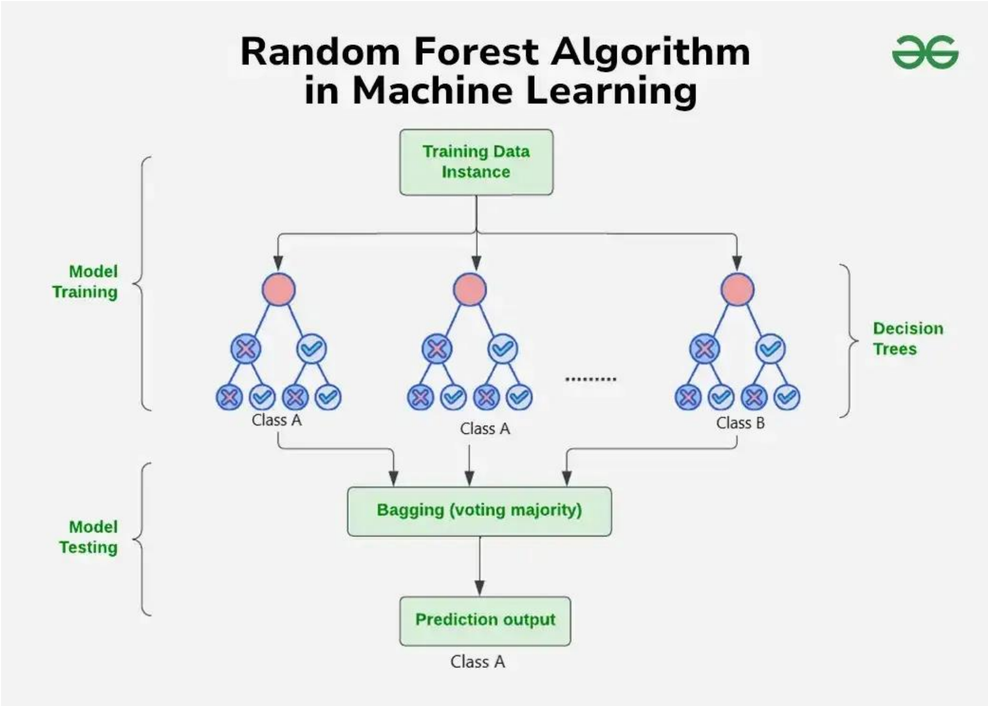


Figure 1: Working of RandomForest Algorithm (Random Forest Algorithm in Machine Learning, 2025)

After generating the prediction, the program returns the estimated value of the target variable based on the input data, and the importance of each of the other variables in the data to further understand the data that the algorithm is working with and the reasoning behind the prediction.

2.1.3 OTHER IMPORTANT PACKAGES

The previously explained packages needed to be further explained as they are crucial to the development of the thesis. However, there are other packages which are key for smaller but yet important tasks:

- GGPlot2, Plotly: GGPlot2 allows the generation of plots, which help visualizing and understanding better the data. Additionally, Plotly package will make those plots more interactive.
- DPLYR, Utils, DT, data.table: These packages will enable more efficient data processing when working with the GIDAS database. DPLYR and Utils help working

with data frames and DT and data.table help turning those data frames into data tables to increase the capabilities of the program, for example, to show the output table in the GUI.

- KeyboardSimulator, clipr: KeyboardSimulator package allows to automate processes by telling the computer to move the mouse, to click it or to press some key on the keyboard. That means that with the right commands, the computer can work on its own for several hours simulating a user. Clipr package allows users to save variables in the clipboard and, therefore, allowing the user to paste them outside the R environment, as well as saving the information in the clipboard in a variable, registering information which is external to the program. These mentioned functions are used for automating the results obtention from PC-Crash.

2.2 GIDAS DATABASE

The German In-Depth Accident Study (GIDAS) is the foundational data source for this thesis, providing a rich and detailed collection of real-world traffic crash data. Established as a collaborative effort between the Federal Highway Research Institute (BASt) and the German Association for Research in Automobile Technology (FAT), GIDAS is designed to support in-depth accident research and vehicle safety analysis (Otte, 2003).

GIDAS employs a statistically representative sampling scheme to ensure that the data collected reflects the broader population of road traffic crashes in Germany. Data is gathered continuously from three urban regions (Dresden, Hanover, and Munich), based on a stratified random sampling method that considers variables such as time of day, weather conditions, and accident severity. This approach ensures that the dataset captures a wide range of crash scenarios, making it suitable for both descriptive and inferential analyses.

The database itself is structured into multiple interrelated tables, each focusing on different aspects of a crash event. These include information on vehicles, road users, environmental conditions, injuries, and technical reconstructions. For this thesis, the most important table is Reko (short for Rekonstruktion), which contains detailed reconstruction data for individual collision sequences. The Reko table includes variables such as:

- Pre-crash speeds and trajectories
- Impact angles and points of contact
- Energy Equivalent Speed (EES)
- Delta-v and relative velocity values

This data is generated through a combination of on-site crash investigation, vehicle inspections, and post-crash reconstruction using simulation tools and expert analysis. The high level of detail in Reko allows for precise modeling of crash dynamics and is essential for evaluating crash severity and vehicle safety performance.

In addition to Reko, other relevant tables will be referenced throughout the thesis as needed. These may include vehicle specifications and deformation measurements, amongst others, all of which contribute to a holistic understanding of each crash event.

2.3 EES ANALYZER

Using R, a user interface will be created by the name EES Analyzer. It will be used as a powerful tool for identifying and checking cases with uncommon values of EES and, overall, getting conclusions easier, thanks to easy access to the database. Those conclusions will contribute to analyze the compatibility between vehicles during a car crash. A more detailed explanation of this tool can be found later in this document, in Chapter 5: Developed System.

2.4 PC-CRASH

PC-Crash is a specialized software used for traffic accident reconstruction. It allows users to simulate and analyze vehicle collisions in 2D and 3D, helping experts determine factors like vehicle speeds, trajectories, impact forces, and crash dynamics. It is widely used by forensic engineers, police, and accident investigators to create technical reports and visual reconstructions. However, for this thesis, the only relevant feature of this software will be the EES-CNN tool.

The EES-CNN tool is a system that uses machine learning to see images and return an estimated EES value. The way this tool functions starts with an image recognition tool, which will analyze the uploaded image, and it will identify where the damage of the crash is and how severe it is. Then, it will compare the uploaded image with an internal repository with a high number of different cases and images and uses it to estimate a range of values for EES, also showing the confidence that the program has.

Chapter 3. STATE OF THE ART

This chapter reviews the current advancements and methodologies in estimating the energy equivalent speed (EES) and the technical assessment of crash severity. It examines the models and algorithms used to improve the accuracy of EES evaluations. Additionally, the challenges and limitations in re-evaluating EES values, including noise and irregularities in the data, are discussed. The review also addresses the importance of considering partner protection in crash assessment models, aiming to minimize damage to all parties involved in traffic incidents. Finally, recent studies that have utilized EES databases to identify factors contributing to inaccurate or irregular EES coding are explored, with the goal of enhancing road safety and reducing traffic fatalities.

3.1 ORIGINS OF THE EES CONCEPT

One of the foundational contributions to the quantification of vehicular collision severity is the work of Campbell (1974), who introduced an energy-based framework for evaluating crash impacts. In his seminal study (Campbell, 1974), Campbell proposed the concept of Equivalent Barrier Speed (EBS) as a standardized metric to represent the severity of real-world automobile collisions. EBS is defined as the speed of the kinetic energy dissipated due to deformation. In other words, when you know the energy dissipated due to deformation of the car crash, and define it as kinetic energy, EBS is the velocity at which the vehicle would carry that exact energy. This approach enables a consistent comparison between controlled crash tests and field accidents.

The methodology is grounded in the analysis of residual crush (the permanent deformation of the vehicle structure) as a proxy for the energy absorbed during impact. By leveraging data from full frontal barrier tests, Campbell developed a linear force-deflection model that relates crush depth to impact speed. This model assumes uniform structural stiffness across the vehicle front and simplifies the estimation of absorbed energy through integration over the damaged area.

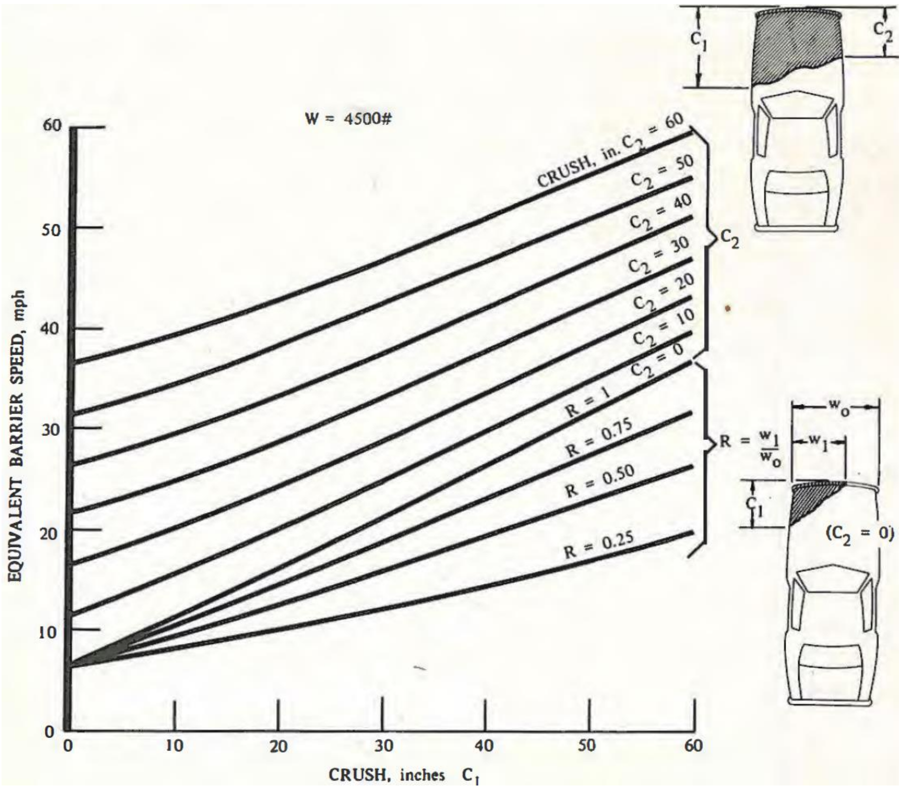


Figure 2: EBS vs crush for '71-'72 full size GM vehicles having angle barrier damage patterns (Campbell, 1974)

To validate the model, Campbell compared EBS estimates with actual impact speeds from controlled angle and offset barrier tests. Figure 2 shows the result of this model for different crash angles. The results demonstrated high accuracy, with average errors within ±3–4 mph, indicating the model's robustness for a range of frontal impact configurations. Furthermore, a pictorial estimation technique was introduced, allowing practitioners to visually approximate EBS by segmenting the vehicle front and summing energy contributions from each section (Figure 3).

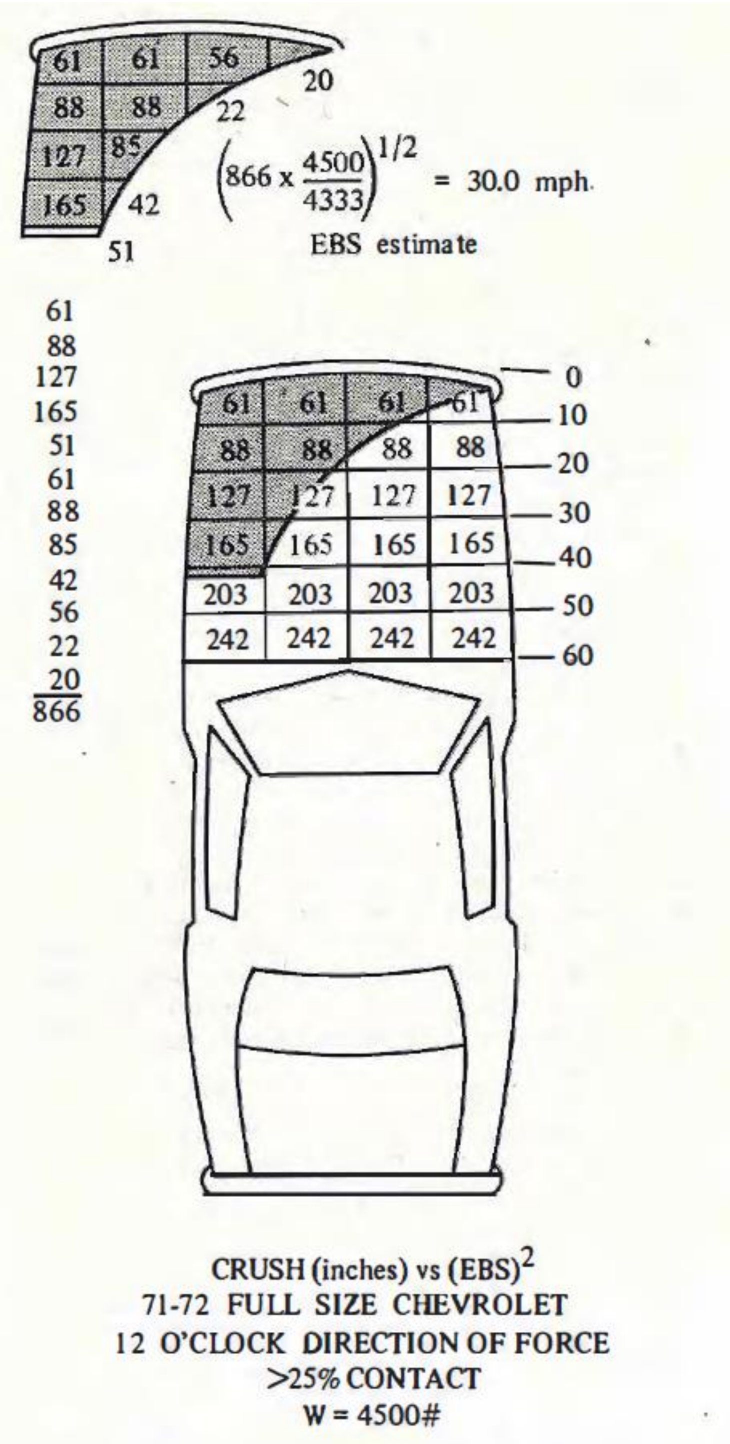


Figure 3: Pictorial approach to 30mph offset barrier impact (Campbell, 1974)

While the model is limited to frontal impacts and assumes uniform vertical damage, its general framework is extensible to side and rear collisions. Importantly, Campbell emphasized that EBS captures only one dimension of collision severity. Other factors, such as impact direction, object stiffness, and duration, also influence injury outcomes and must be considered when evaluating occupant protection systems.

Campbell's work laid the groundwork for subsequent research in crash severity analysis and remains a cornerstone in the development of objective, energy-based crash metrics. His approach provides a critical link between vehicle deformation and occupant injury potential, facilitating more accurate assessments of safety system performance in real-world conditions.

3.2 DEVELOPMENT OF THE COMPATIBILITY CONCEPT

The concept of compatibility in passenger vehicle safety represents a critical intersection of engineering, accident analysis, and public health. It refers to the ability of vehicles to interact in a crash in a way that minimizes injuries and fatalities for all parties involved. This includes both self-protection, which is how well a vehicle protects its own occupants, and partner protection, which is how much harm it causes to occupants of the other vehicle. Robert Zobel's 1998 paper provides a comprehensive examination of this issue, emphasizing the need for a balanced approach that considers the structural behavior of vehicles in real-world collisions, not just in controlled crash tests (Zobel, 1998).

Vehicle compatibility becomes particularly relevant in car-to-car collisions, where differences in mass, stiffness, and structural geometry can lead to disproportionate injury outcomes. Larger, stiffer vehicles often inflict greater damage on smaller, lighter ones, a phenomenon referred to as aggressiveness. Conversely, a vehicle's ability to minimize harm to its collision partner is termed partner protection. The challenge lies in designing vehicles that achieve both high self-protection and high partner protection, a task complicated by the diversity of vehicle types on the road. Figure 4 shows the size of the opposing car in an accident when the driver of the first vehicle was injured, and it can be seen that bigger vehicles are more present in AIS3+ or AIS4+ than lighter vehicles, which means that lighter vehicles are better for partner protection than heavier vehicles.

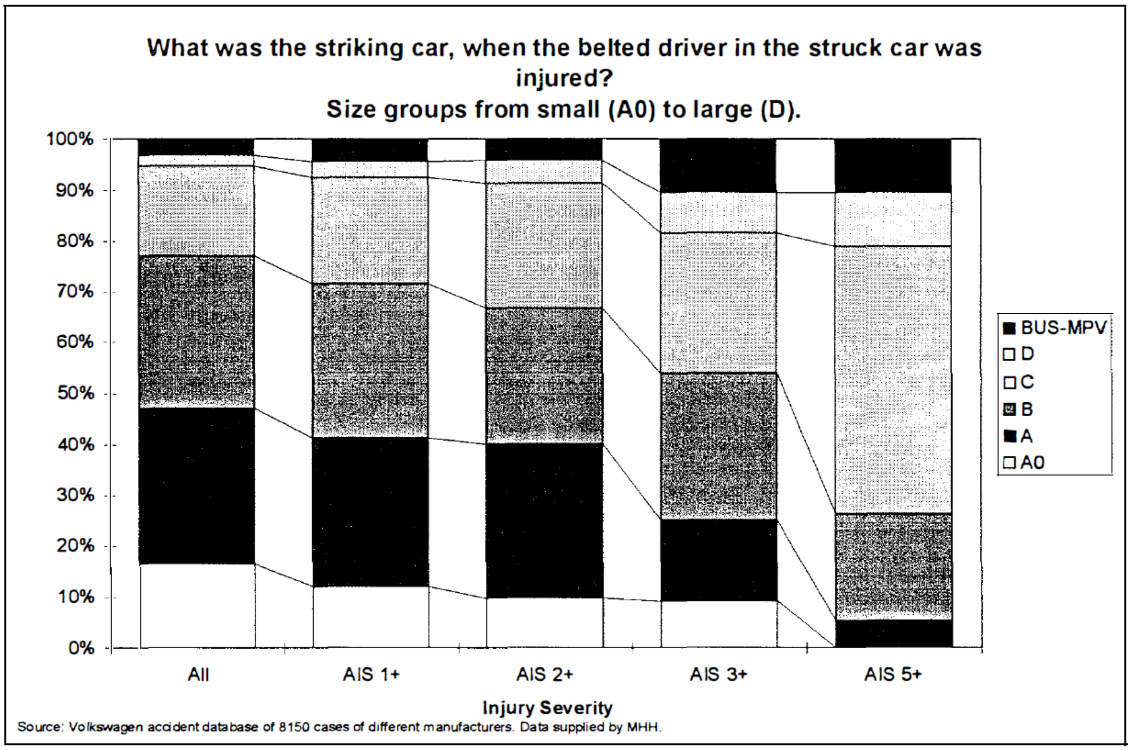


Figure 4: Hazard of different size groups crashes between passenger cars (Zobel, 1998)

Zobel highlights that while crash tests provide valuable insights into vehicle deformation patterns, they do not always correlate with real-world injury outcomes. For instance, a vehicle that deforms more in a crash test may not necessarily result in higher occupant injuries, and vice versa. Therefore, accident data analysis is essential to validate the relevance of crash test observations. This requires detailed databases that capture not only vehicle and occupant parameters but also structural deformation characteristics, which are often difficult and costly to obtain.

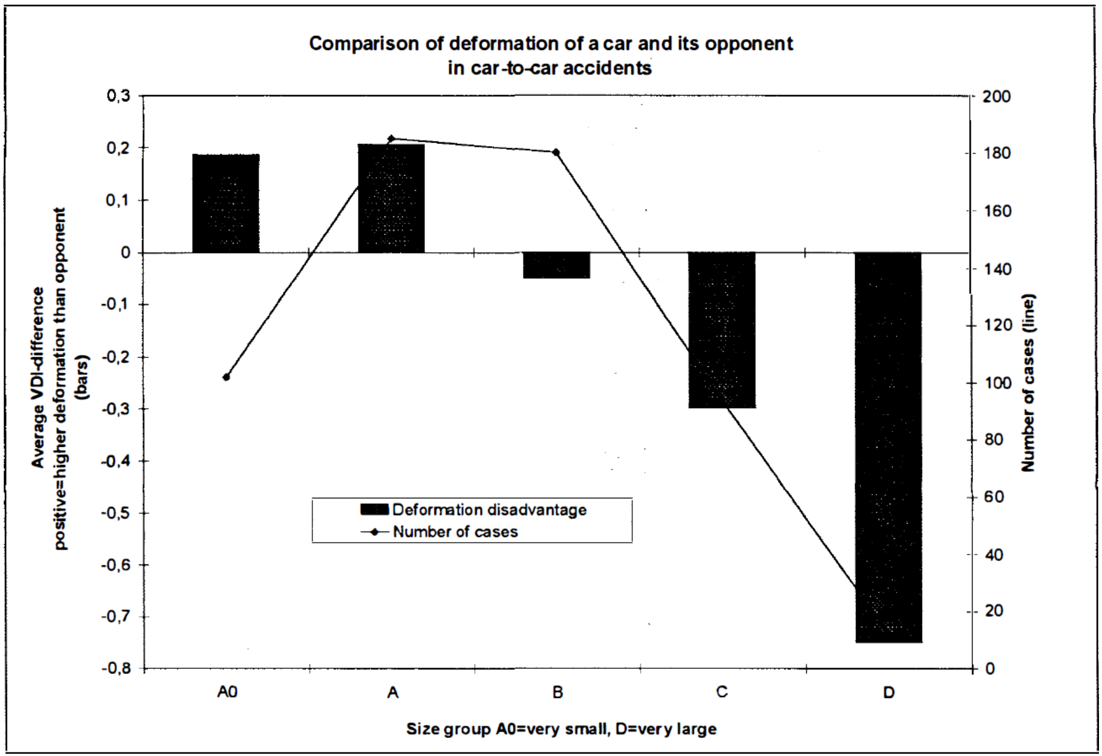


Figure 5: Deformation behavior of different size groups, based on a comparison of VDI 6 in car-to-car crashes (Zobel, 1998)

The paper identifies several key characteristics that influence compatibility. Vehicle mass is the most dominant factor, as it directly affects the velocity change experienced by the struck vehicle, which in turn influences injury risk. Figure 5 shows how cars suffer against each group size. The figure states how the deformation suffered is lower than the opposing vehicle when the opposing vehicle is the same size or smaller and has a higher value for bigger opposing vehicles.

Vehicle stiffness also plays a role, though its impact is less clear-cut. Other factors include the height and geometry of the vehicle's front end, the orientation of the engine (longitudinal vs. transverse), and the distribution of structural forces during impact. For example, vehicles with higher front ends tend to offer less partner protection in side impacts, while a well-balanced force distribution in the front structure may enhance compatibility (Zobel, 1998).

Another concept introduced in the paper is the "bulkhead concept." This approach proposes a structural design that limits the maximum force a vehicle can exert during a collision. By incorporating a bulkhead that caps force levels, vehicles can avoid excessive intrusion into the occupant compartment, even in high mass-ratio collisions. The theoretical foundation of this concept is based on energy conservation principles: if two vehicles are designed for a certain barrier impact speed, and their closing speed in a collision does not exceed twice that speed, then sufficient deformation capacity exists to prevent compartment collapse. This holds true regardless of the mass ratio between the vehicles, provided both are designed to deform appropriately.

However, the bulkhead concept is not without limitations. Practical constraints such as vehicle length, acceptable deceleration levels for occupants, and the capabilities of restraint systems must be considered. For instance, a deceleration of 30 g is already at the upper limit of what restraint systems could manage without causing injury, particularly for older occupants (Zobel, 1998). Additionally, the available deformation stroke in a vehicle's front structure is limited by design and packaging constraints. These factors restrict the applicability of the bulkhead concept to a certain range of mass ratios. In the German vehicle fleet, these ratios can go up to 1.6, which still covers approximately 90% of real-world frontal collisions.

Zobel's work represents a pivotal shift toward system-level thinking in vehicle safety, advocating for compatibility as a fleet-wide property rather than an attribute of individual vehicles. His findings have laid the groundwork for ongoing research and policy development in both Europe and North America, emphasizing the importance of integrated, data-driven approaches to enhance crash outcomes for all road users.

3.3 CRASH CONFIGURATION TYPOLOGIES

Accurate classification of crash configurations is essential for understanding injury mechanisms and evaluating vehicle safety performance. Brumbelow (2019) introduced a refined typology of frontal crash configurations based on photographic review of real-world collisions (Brumbelow, 2019). These configurations are defined relative to the engagement of a vehicle's primary longitudinal structures and the direction of force application.

The study identified seven distinct configurations, which are the following:

- Large Overlap
- Moderate Overlap
- Small Overlap
- Center Impact
- Perpendicular
- Oblique Center
- Oblique Corner

A more detailed explanation is shown in ANEX I. Even though Table 7 shows eight configurations, override is not considered a crash configuration. It is used more as an exclusion criterion, as the longitudinal structures cannot be loaded properly, or to compare the damage between underride and non-underride car crashes.

This classification enables a more precise correlation between crash geometry and injury risk. It also highlights the limitations of traditional delta-V estimates in non-standard impact scenarios, reinforcing the need for configuration-specific analysis in crashworthiness research.

3.4 RECENT ADVANCES IN EES CALCULATION

While previous studies have laid important groundwork, a more recent investigation, conducted by Pascal Breitlauch in 2023, proposed a novel method for calculating the EES in traffic crashes (Breitlauch, Junge, Erbsmehl, Sandner, & van Ratingen, 2023). Accurate and objective quantification of crash severity remains a central challenge in traffic safety research and accident reconstruction. Traditional methods often rely on subjective assessments or expert estimations, which can introduce significant bias and variability (Meghna Chakraborty, 2023). In response to this limitation, Breitlauch proposed a novel, model-based approach for estimating crash severity by calculating the EES from the visual deformation of the vehicle. While EES itself is an established metric in crash analysis, his contribution lies in developing a data-driven methodology that links observable post-crash deformation patterns to EES values, enabling more consistent and automated severity assessments.

The core of the methodology is a voxel-based model that maps post-crash deformation data onto a three-dimensional representation of the vehicle. Vehicles are categorized into six types (e.g., super-compact, sedan, SUV), and deformation data from real-world crash databases (GIDAS and NASS CDS) are used to construct intermediate EES models. These models are then stratified by Principle Direction of Force (PDOF) and Vehicle Type (VT) to create structure EES models, which are subsequently normalized using standardized crash test data from EuroNCAP and ADAC.

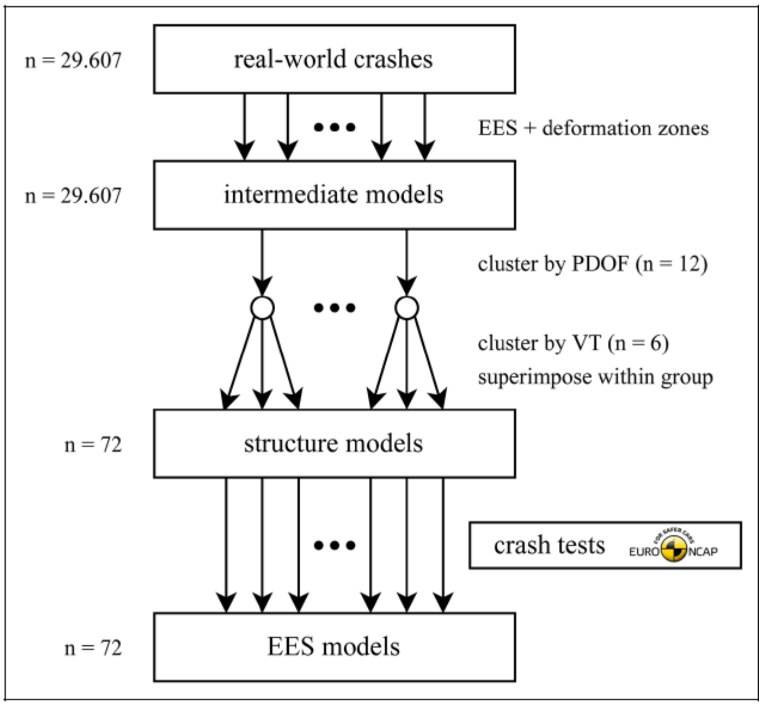


Figure 6: Diagrammatic representation of the developed method (Breitlauch, Junge, Erbsmehl, Sandner, & van Ratingen, 2023)

A key innovation of this approach is the use of crash test normalization to eliminate subjective bias from real-world data. By comparing deformation energy from crash tests with known EES values, the model adjusts the structure EES outputs to reflect objective crash severity. The final EES models can then be applied retrospectively to crash databases, enabling consistent and unbiased severity estimation across large datasets.

Validation of the model was performed using a controlled car-to-car crash test between an Audi Q7 and a Fiat 500. The EES values computed by the model closely matched those measured during the test, with deviations of only 2.5% and 6.5%, respectively. Additionally, application of the model to the GIDAS database revealed systematic biases in traditional EES estimates, particularly overestimation in frontal and rear-end collisions and underestimation in side impacts.

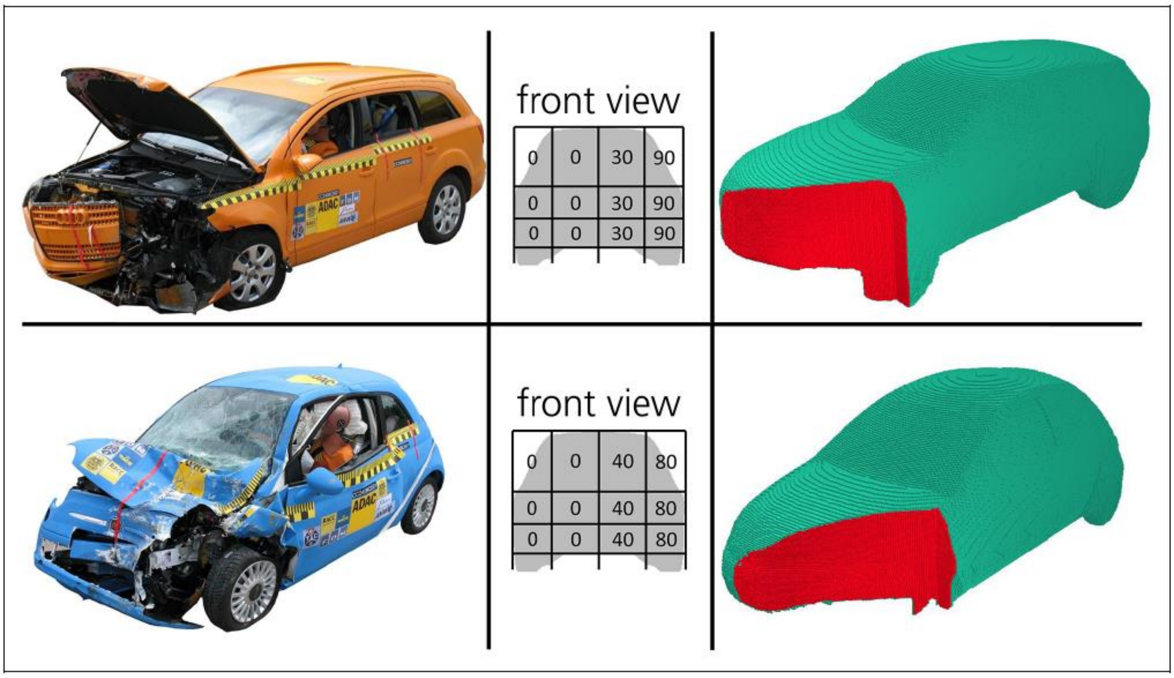


Figure 7: Audi Q7 vs Fiat 500. 50% overlap car-to-car crash test (Breitlauch, Junge, Erbsmehl, Sandner, & van Ratingen, 2023)

Breitlauch et al.'s work represents a significant advancement in the field of crash analysis by providing a scalable, objective, and reproducible method for estimating crash severity. The integration of voxel modeling, real-world crash data, and standardized testing creates a robust framework that can enhance the reliability of crash databases and support more accurate evaluations of vehicle safety performance.

All the previously mentioned studies, along with many other relevant research efforts, share the ultimate goal of achieving an objective and reliable way to calculate EES in any traffic crash. This objective aligns with the aim of this thesis, which seeks to enhance road traffic safety by developing a comprehensive database for evaluating crash severity using real-world crash data. By improving the accuracy of technical crash severity ratings and reducing noise and irregularities in EES assessments, this thesis contributes to the ongoing efforts to reduce traffic fatalities and enhance overall road safety.

3.5 VEHICLE COMPATIBILITY IN MODERN FLEETS

The issue of vehicle incompatibility, where disparities in vehicle mass and design lead to unequal crash outcomes, has long been a central concern in traffic safety research. Historically, heavier vehicles such as SUVs and pickup trucks have been shown to offer superior protection to their own occupants while posing elevated risks to occupants of lighter vehicles in multi-vehicle collisions (Gabler & Hollowell, 1998; Eric R & Nolan, 2012). Recent empirical work by Monfort (2024) provides a comprehensive update on this phenomenon, analyzing crash data from 2011 to 2022 to assess trends in vehicle aggressiveness and self-protection across the U.S. passenger vehicle fleet.

Monfort's findings indicate a notable improvement in crash compatibility between cars and larger vehicles in the most recent period (2017–2022), particularly among the heaviest SUVs and pickups. For example, pickups that were previously 2.5 times more likely than cars to fatally injure a car driver in a crash were only 1.9 times more likely in the later period. Similarly, SUVs over 2.250kg reduced their relative aggressivity from 1.9 to 1.2 times that of cars. These improvements are attributed to structural design changes, such as the alignment of energy-absorbing structures, and the proliferation of advanced safety technologies like automatic emergency breaking and side-curtain airbags. The figure below shows the improvement in crash compatibility over time, especially for pickups and SUVs.

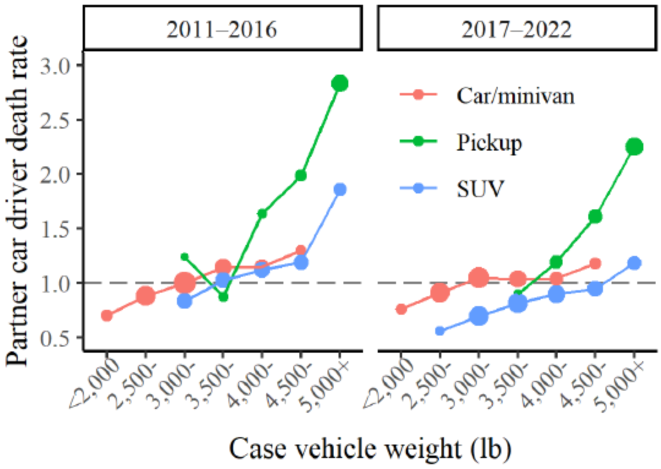


Figure 8: Aggressiveness by Vehicle Type and Weight (Monfort, 2025)

However, Monfort also highlights a critical trade-off: while increased curb weight enhances self-protection for lighter vehicles (especially those under 2.000 kg), it offers diminishing returns for heavier vehicles and significantly increases the risk to crash partners. The figure below illustrates the self-protection benefit of curb weight, showing that protection plateaus around 2.000 kg, meaning that at some point there is no benefit in heavier vehicles.

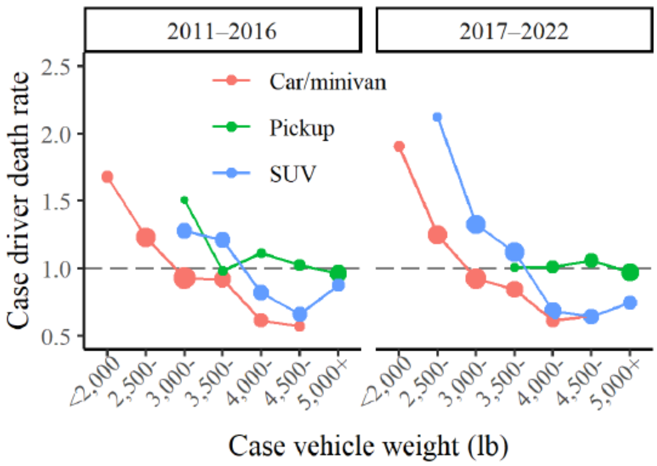


Figure 9: Self Protection by Vehicle Weight (Monfort, 2025)

This asymmetry suggests that the societal benefit of additional vehicle mass is not linear and may, in fact, become negative beyond a certain threshold. Figure 10 shows how lighter vehicles benefit more from added mass, while heavier ones impose greater risk.

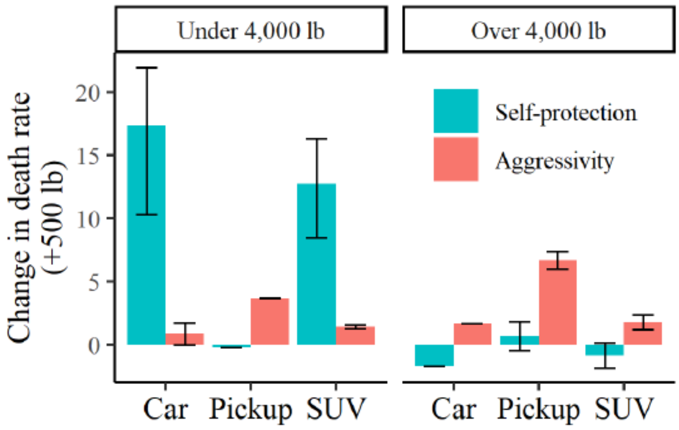


Figure 10: Weight Trade-Off: Protection vs Risk (Monfort, 2025)

A similar conclusion had already been drawn and expanded upon in a 2024 investigative report by The Economist, which analyzed over 7.5 million two-vehicle crashes across 14 U.S. states (The Economist, 2024). Their analysis found that the heaviest 1% of vehicles (around 3,000 kg) were responsible for an average of 37 partner-vehicle deaths per 10,000 crashes, more than six times the rate for median-weight vehicles and over 14 times that of the lightest 1%. While these heavy vehicles experienced fewer fatalities among their own occupants, the net societal cost was stark: for every life saved inside a heavy SUV or truck, more than a dozen were lost in other vehicles.

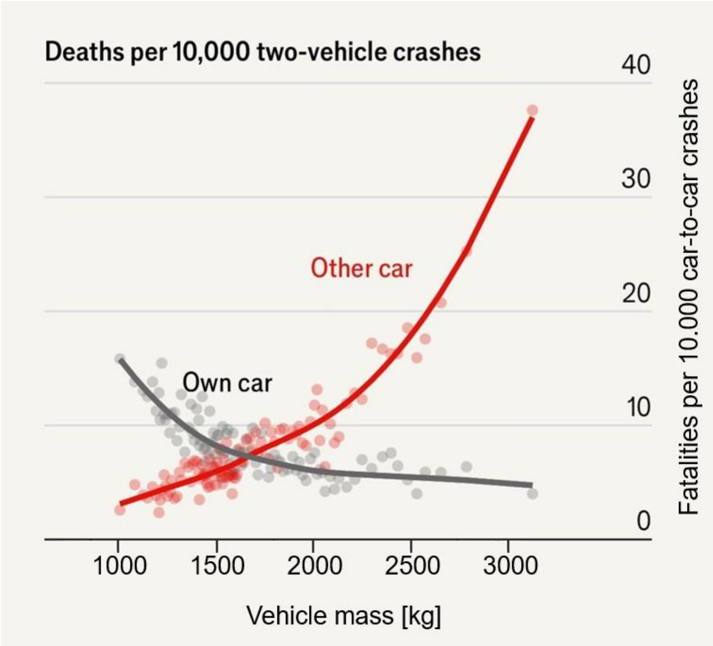


Figure 11: Fatalities by Vehicle Weight (The Economist, 2024)

Moreover, The Economist quantified the potential impact of downsizing the heaviest vehicles. If the top 10% of vehicles by weight (those over 2.250 kg) were reduced to the next-lower weight class (2.000 – 2.250 kg), the U.S. could see a 12% reduction in multi-vehicle crash fatalities (equivalent to approximately 2,300 lives saved annually) without compromising the safety of the heavier vehicles' occupants.

Both Monfort and The Economist converge on a critical insight: the protective benefits of vehicle mass plateau at around 1.800 – 2.000 kg, while the externalized risks to other road users continue to rise (Monfort, 2025; The Economist, 2024). This creates a compelling case for policy interventions aimed at curbing the proliferation of excessively heavy vehicles. Yet, as The Economist notes, market trends are moving in the opposite direction. In 2023, vehicles over 2250 kg accounted for 31% of new U.S. vehicle sales, up from 22% just five years earlier. This growth is fueled not only by consumer preferences but also by regulatory incentives, such as lenient fuel-efficiency standards for light trucks and tax deductions for heavy-duty business vehicles.

In sum, the current body of evidence underscores a growing misalignment between individual safety gains and collective road safety outcomes. While technological advancements have mitigated some aspects of vehicle incompatibility, the unchecked increase in vehicle mass, particularly among SUVs and pickups, continues to impose a disproportionate burden on other road users. Future research and policy must address this imbalance, potentially through mass reduction strategies, revised crash testing protocols that account for inter-vehicle harm, and regulatory reforms that disincentivize excessive vehicle weight (Insurance Institute for Highway Safety (IIHS), 2021).

Chapter 4. DEFINITION OF THE WORK

In this chapter, it will be explained the reason this project is important. There will also be shown all the different objectives that this project aims to fulfill, and the methodology followed to reach those goals.

4.1 JUSTIFICATION

Despite decades of progress in vehicle safety engineering and crash testing protocols, road traffic crashes remain a leading cause of death and injury worldwide (World Health Organization, 2024).

Current crashworthiness assessments, conducted by Euro NCAP or IIHS, have significantly improved vehicle design. These are based on the evaluation of how well vehicles protect their occupants in standardized crash scenarios. They focus primarily on injury risk to the vehicle's own occupant, and they often rely on standardized test conditions (European New Car Assessment Programe (Euro NCAP), 2017; Insurance Institute for Highway Safety (IIHS), 2021).

However, this disconnect becomes particularly relevant when estimating crash severity using metrics like Energy Equivalent Speed (EES), especially in retrospective crash databases such as GIDAS. In these cases, EES values are often reconstructed or inferred based on incomplete or idealized assumptions, which may not accurately reflect the actual crash dynamics. Factors such as multi-impact sequences, oblique angles, vehicle mismatches, and post-impact behavior introduce uncertainties that standardized models struggle to account for. Therefore, while discrepancies between crash tests and real-world conditions are a contributing factor, the issue of inaccurate EES codings in GIDAS also stems from methodological limitations in crash reconstruction and data interpretation.

The proposed project addresses this critical shortcoming by developing a comprehensive, data-driven crash severity database that re-evaluates and enhances the accuracy of EES coding using real-world crash data. Unlike traditional approaches that depend heavily on simplified assumptions or static deformation measurements, this project leverages validated crash configurations, vehicle-specific parameters, and high-fidelity datasets such as GIDAS to recalibrate EES values with greater objectivity and technical rigor.

The primary objective of this project is to support crash research by developing a tool that enables the retrospective estimation of Energy Equivalent Speed (EES) based on scene photographs and vehicle-specific parameters. The motivation stems from the need to validate and, where appropriate, improve existing EES codings within the GIDAS database. These codings are often derived under conditions of limited information and may not fully capture the technical severity of a crash. By refining EES estimations, the tool aims to enhance the accuracy of crash severity assessments, which in turn can contribute to more robust analyses of partner protection and mass ratio effects in mixed-vehicle collisions. While the tool is primarily intended for internal research use, it may also offer methodological value to other users of in-depth crash data.

In an era where data-driven safety solutions are increasingly prioritized, this project stands out as a timely and impactful contribution. It not only enhances the scientific understanding of crash dynamics but also provides a practical foundation for improving road safety outcomes through better-informed engineering and regulatory practices.

4.2 OBJECTIVES

The evaluation of crash severity is a cornerstone of modern road safety research and vehicle safety engineering. Accurate severity assessments are essential for understanding injury mechanisms, improving vehicle design, and informing regulatory standards. However, current methodologies often fall short in capturing the complexity of real-world crashes, particularly in the estimation of Energy Equivalent Speed (EES). This project addresses these limitations through a set of well-defined objectives aimed at developing a robust, data-driven framework for crash severity analysis. The following subsections outline the specific goals that will guide the project's development and implementation.

4.2.1 Develop a Comprehensive Crash Severity Database

The first objective is to construct a high-resolution, multi-dimensional database that consolidates real-world crash data from GIDAS. While the raw values, such as vehicle types, crash configurations, deformation patterns, delta-V values, and injury outcomes, are already available within GIDAS, the novelty of this database lies in its structured and stratified design. By organizing the data in a way that enables direct comparisons between similar vehicle models involved in similar crash configurations, the database facilitates the identification of outliers and inconsistencies against the prediction of the model. This structure not only supports more meaningful analyses of crash severity but also helps improve data coding quality by highlighting anomalies. Additionally, the database is designed for scalability and interoperability, allowing for future integration with simulation tools and machine learning models.

4.2.2 Improve Accuracy of EES Codings

The second objective focuses on enhancing the reliability of EES. Although it is a well-established concept in traffic crash reconstruction, its practical application often suffers from significant limitations due to the way it is typically estimated in real-world crash analysis.

In many cases, EES values are derived from visual inspections of post-crash vehicle deformation, a process that introduces a high degree of subjectivity and uncertainty. Analysts typically rely on photographs, sketches, or on-site assessments to estimate crush depth and distribution, which are then used to infer the energy absorbed during the collision. Crucially, the accuracy of these estimations depends heavily on the competence and experience of the individual analyst. Misinterpretation of deformation patterns overlooked internal structural damage, or failure to account for asymmetries in the crash configuration can all lead to significant deviations in the resulting EES values. This reliance on expert judgment, while often necessary, underscores the need for tools that can support more consistent and objective assessments.

These visual estimations are further complicated by the diversity of modern vehicle designs. Differences in materials, structural reinforcements, and energy absorption strategies mean that similar-looking deformations can correspond to vastly different crash severities. As a result, assigning an accurate EES value becomes particularly challenging when comparing vehicle models or crash types.

To address these issues, this thesis proposes a systematic re-evaluation of EES values using validated crash reconstructions and a stratified database structure. By grouping similar vehicle models and crash configurations, the methodology enables the identification of potential outliers and inconsistencies in EES estimation. This comparative approach not only helps to refine the accuracy of severity ratings but also supports the improvement of data coding practices and the development of more objective, reproducible estimation techniques.

4.2.3 Identify Contributing Factors to Inaccurate EES Coding

A critical analytical goal is to investigate the root causes of discrepancies in EES coding. Using the developed database, the project will analyze how factors such as crash configuration, vehicle mass ratio, structural misalignment, and deformation location contribute to irregular or misleading EES values. This analysis will help identify patterns of error and inform the refinement of crash reconstruction methodologies. The findings will be particularly valuable for improving the accuracy of automated crash analysis tools and for guiding future updates to crash severity coding standards.

4.2.4 Contribute to the Reduction of Traffic Fatalities

The overarching objective of this project is to contribute to the broader effort of reducing road traffic injuries and fatalities by improving the accuracy and objectivity of crash severity assessments. More reliable estimations of EES can enhance the evaluation of vehicle safety performance and support evidence-based policy and design decisions. A key application of this work lies in the analysis of partner protection, where EES serves as a core metric for identifying severity discrepancies between vehicles of different sizes and masses. By enabling more consistent assessments of crash severity, particularly in mixed-vehicle collisions, the project supports a deeper understanding of mass ratio effects and their implications for occupant safety.

4.3 METHODOLOGY

The methodology for this project is designed to systematically address the challenge of accurately evaluating crash severity using real-world data. It combines data engineering, statistical modeling, and user interface development to create a robust and accessible system for analyzing Energy Equivalent Speed (EES) and related crash metrics. The approach is divided into three main phases: data collection and preparation, data analysis and model development, and visual estimation automation.

4.3.1 DATA COLLECTION AND PREPARATION

4.3.1.1 Creating a New Database Using R:

The first step involves extracting and structuring relevant data from the GIDAS database. Using the R programming language within the RStudio environment, a new, purpose-built database is created to store and manage crash data efficiently. This process includes:

- Filtering for relevant crash types (e.g., car-to-car frontal impacts).
- Selecting key variables such as vehicle type, crash configuration, deformation measurements, and injury outcomes.
- Cleaning and transforming the data to ensure consistency and usability.
- Structuring the data into a relational format that supports advanced querying and analysis.

This curated dataset forms the foundation for all subsequent analyses and model development.

4.3.1.2 Developing a Graphical User Interface (GUI):

To facilitate user interaction with the database, a graphical user interface (GUI) is developed. The GUI allows users to:

- Browse and visualize individual crash cases.
- Apply filters based on crash configuration, vehicle type, or severity metrics.
- Export selected data for further analysis.

The GUI is implemented using Shiny for R, ensuring accessibility for both technical and non-technical users. This interface enhances the usability of the database and supports exploratory data analysis.

4.3.2 DATA ANALYSIS AND MODEL DEVELOPMENT

4.3.2.1 Re-evaluating Existing EES Values:

A core component of the project is the re-evaluation of existing EES values to improve their accuracy. This involves:

- Comparing cases with the same crash configuration (e.g., moderate overlap, small overlap, underride).
- Using statistical techniques such as linear regression and residual analysis to quantify and reduce estimation errors.

The goal is to align crush-based EES values with data-based EES values, thereby enhancing the objectivity and consistency of crash severity assessments.

4.3.2.2 Identifying Contributing Factors to Inaccurate EES Coding:

To further improve the reliability of EES as a crash severity metric, the project investigates the root causes of inaccurate or inconsistent EES coding. This analysis includes:

- Using regression models and pattern recognition to identify correlations between crash characteristics (e.g., impact angle, vehicle mass ratio) and EES discrepancies.

- Categorizing errors by crash configuration, vehicle type, age of the cars, crash year, regions, velocities, etc., to understand where traditional methods fail.

The insights gained from this analysis will provide recommendations for improving EES coding protocols and crash reconstruction methodologies.

4.3.3 PREDICTION TOOL AND VISUAL ESTIMATION AUTOMATION

In order to obtain a corrected value for EES in real world crashes, two different and complementary techniques will be applied.

- A machine learning predictive model that predicts an EES value depending on other variables of the car crash, such as the relative velocity or the deformation.
- An automated method to show images to an external software that uses machine learning to estimate the EES value of the car crash, comparing them to other images integrated in its database.

These two methods will allow the identification and correction of inaccurate EES coding.

4.4 SCHEDULE

The project was divided into three main sections. First, the EES Analyzer tool is developed in depth, as it is the main tool of this project and the basis for the whole study. Once the tool works, and access to PC Crash is enabled, the EES Estimator tool is developed. Finally, both tools are used to obtain results and reach conclusions. The report is written parallel to the rest of the tasks to keep track of each advancement made. Figure 12 shows the timeline of the project, showing in blue the tasks related to EES Analyzer, in green the tasks related to EES Estimator and in orange the tasks needed to obtain results and author the report.

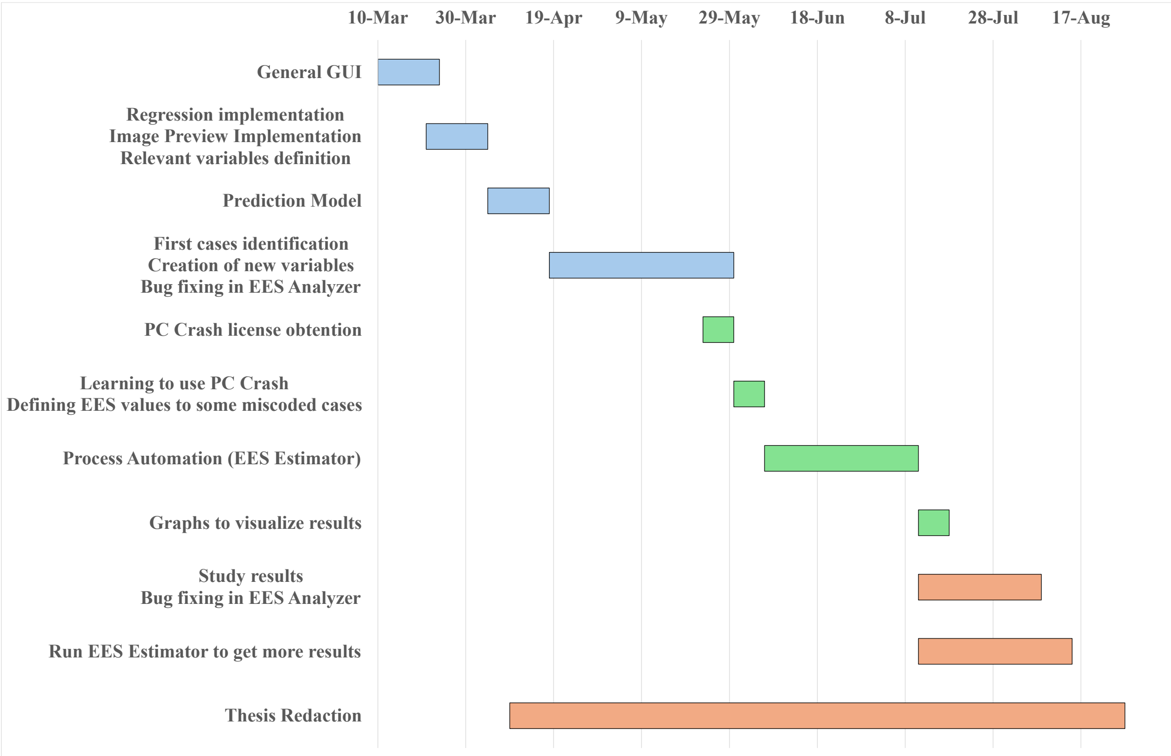


Figure 12: Timeline of the Project

Chapter 5. DEVELOPED SYSTEM

For the analysis of the GIDAS database, a new tool has been created that enables easy comparison of the existing data, easy search of similar cases and analysis of the variables present in the database. In addition, this tool is also able to predict which value of EES each case should have in each case depending on other selected variables.

5.1 DEVELOPMENT OF THE GUI

Using the Shiny library in R, a GUI was generated that allows the user to search for the cases which have a particular crash configuration, a specific car model or a defined car category, which enables the user to search for similar cases and compare them to get conclusions. The user is also able to filter a range of ees values, in case it is intended to obtain only crashes within a range of severity. There is also a checkbox that allows the user to apply the VDI filters. As the crash configuration coding according to (Brumbelow, 2019) is not available for all the cases, there are a great amount of cases that are not classified in any crash configuration. VDI filters analyze the area of the vehicle where the damage has been taken and converts the uncoded value of Brumbelow into a list of possible Brumbelow values, so they are not removed when trying to reach a specific car configuration.

The VDI filters are based on GIDAS codings for VDI1, VDI2 and VDI3. VDI1 defines the principal direction of the force that received the damage. This variable helps identifying whether the crash has a completely frontal impact, has an angle, or is a side or a rare impact. VDI2 defines the main deformed area of the vehicle. In most of the cases, this value is directly with VDI1, as it will also define whether the crash is frontal, side or rare. VDI3 codes the specific horizontal location of the damage, and they depend on VDI2. This last variable helps differentiating distinct types of frontal crashes, as it shows the position of the damage and its width.

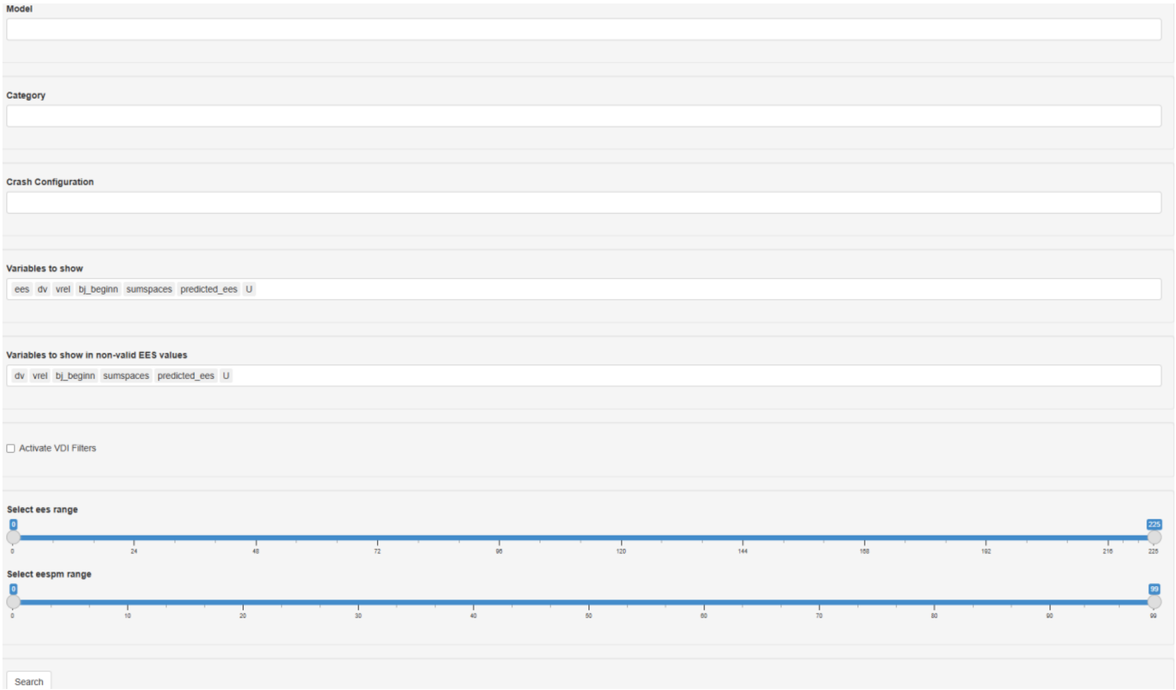


Figure 13: Input Tab of EES Analyzer

After clicking the "Search" button, the UI will offer a table that shows all the cases that fulfill the characteristics asked for by the user, which have an EES value coded. In case the EES value is marked as unknown, it will appear in the table below, where the unknown EES cases are shown. The EES range filter will also be showed in this new window, alongside new checkboxes that allow better filtering capabilities.

One of these filters, "Show only car-to-car crashes," excludes cases involving other vehicle types such as motorcycles or trucks. Another filter, "Show only singular collisions," removes cases involving complex sequences of multiple impacts. This filter was implemented not to imply that multi-collision cases are inherently less severe, but because accurately distributing deformation across multiple impacts is extremely challenging. In many multi-collision scenarios, it is unclear how much of the total deformation corresponds to each individual impact, especially when the sequence involves varying angles, speeds, and contact points. This ambiguity can significantly distort EES estimation and compromise the reliability of severity assessments. By focusing on singular collisions, where the deformation can be more confidently attributed to a single event, the analysis gains clarity and consistency. However, it is important to note that this filter does not imply a judgment on the overall severity of multi-collision crashes but rather reflects a methodological choice to ensure data quality and interpretability.

The information the user wants to see is displayed as a data table in the center of the screen. The user can select different rows from the table, and a preview of the corresponding vehicle appears on the right side of the screen. Additionally, the user has the option to click "Open Directory," which opens the folder associated with each selected case, providing access to all the images that GIDAS has for that specific crash.

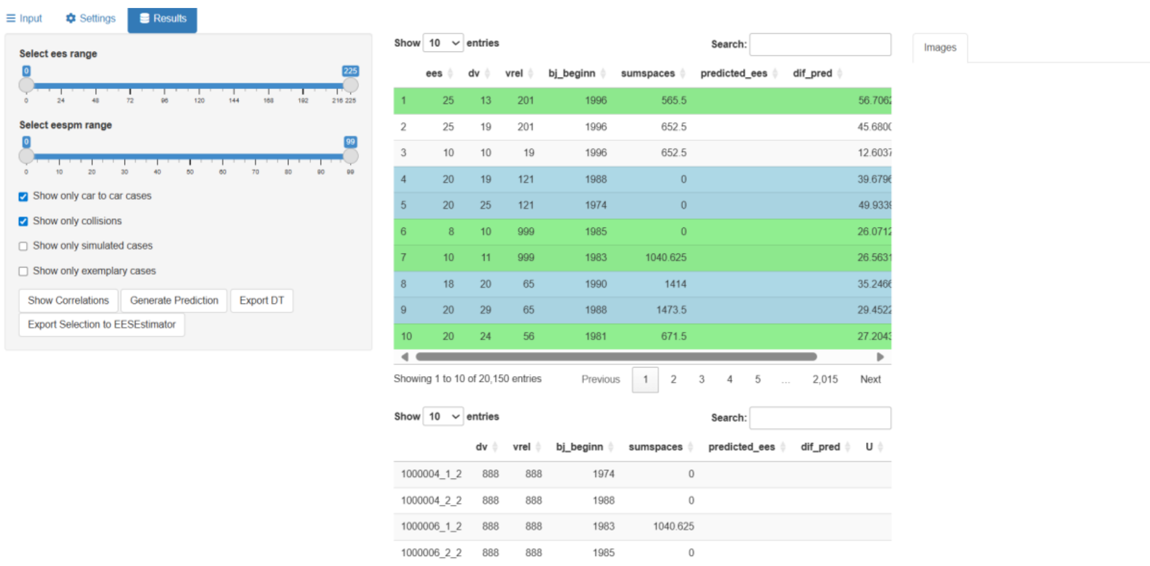


Figure 14: Output Tab of EES Analyzer

In order to find out which variables are relevant, a linear regression can be created between EES and whichever variable the user wants to compare it with. Clicking the "Show Correlation" button, a new window will open showing two selection cells to choose the variables to correlate. Once selected, the user can click on "Run Regression" to receive an output that shows the summary of the linear model, the four correlation plots and a plot that relates the two variables studied.

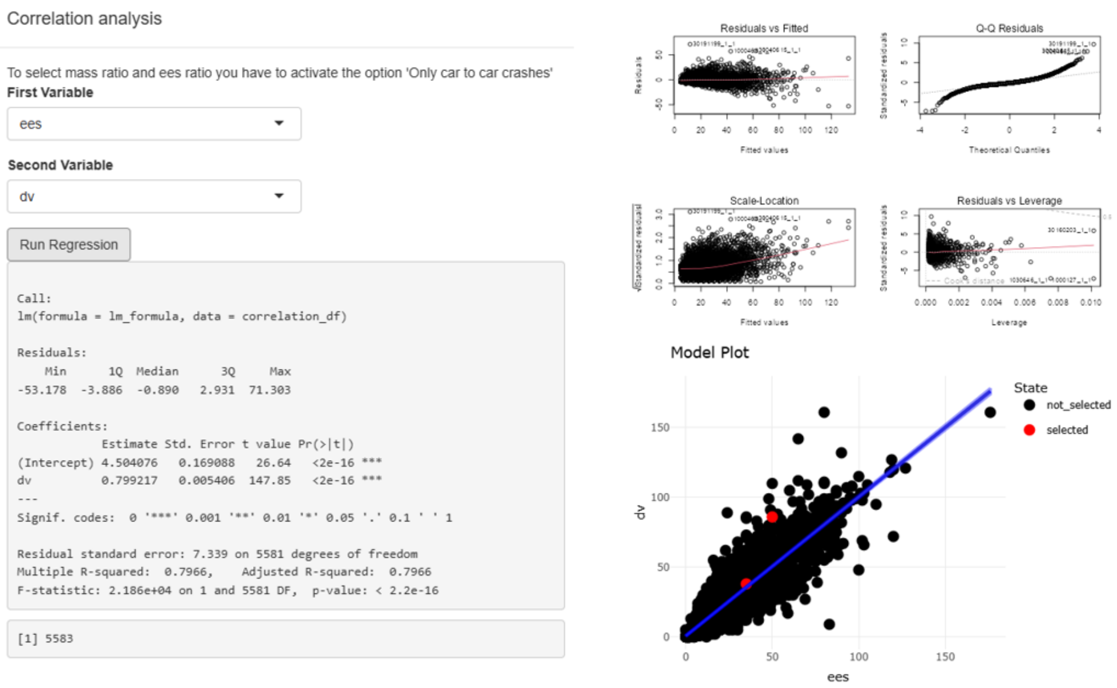


Figure 15: Correlation Analysis Window in EES Analyzer

Once the relevant variables have been determined, the prediction model can be developed. The program has this model previously coded, so the user does not need to use the information from the correlations to generate any prediction. From the user perspective, he will only have to click on “Generate Prediction” and, in the window that appears on the screen, press predict. The program will start generating a model using RandomForest and then using that model to generate a prediction for each variable that appears on both output data tables, as long as the program is able to do so. These predicted variables will be shown in a plot compared to real EES values in the GIDAS, along with other relevant information about the prediction, such as the number of variables the program was able to predict or the precision of the model.

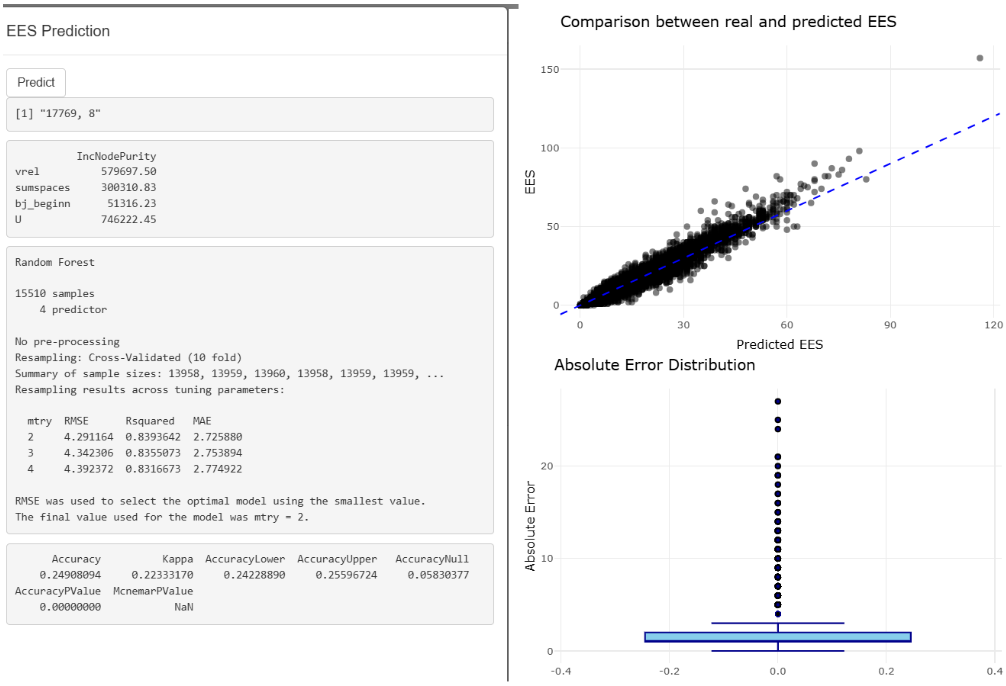


Figure 16: Prediction Generation Window in EES Analyzer

The objective of this program, therefore, is to analyze all frontal car-to-car crashes without severe multiple collisions, with the possibility to do a clustered analysis separating each crash configuration or vehicle classification, use the correlation and prediction tools to identify possible miscoded cases and then compare them to similar car crashes to get conclusions. Moreover, the program has the potential to escalate and analyze all traffic crashes.

The software also allows the user to export the filtered data in two separate ways. By clicking on the "Export DT" button, an Excel file is generated with all the cases shown in the table and the variables selected in the input tab. This tool is useful to generate databases that only contain the specific information that the user is searching for. The second way is the "Export Selection to EESEstimator". This button's main use is to generate a manual input for some cases to the EES Estimator tool, directly from the EES Analyzer. The EES Estimator tool is explained more in depth in: 6.1. EES Estimator.

5.2 *OBTAINING NEW RELEVANT VARIABLES*

GIDAS is a platform that saves a lot of information about each car crash, however, there is additional information that can be obtained from its variables and yet is not a variable itself.

5.2.1 DEFORMATION

The first variable to be obtained is the estimate of the total area of deformation of the vehicle. GIDAS uses four areas of deformation to specify how the vehicle was deformed. However, there is no variable of general deformation, which is the variable that this project intends to study. That means that this variable needs to be calculated in order to perform the analysis intended.

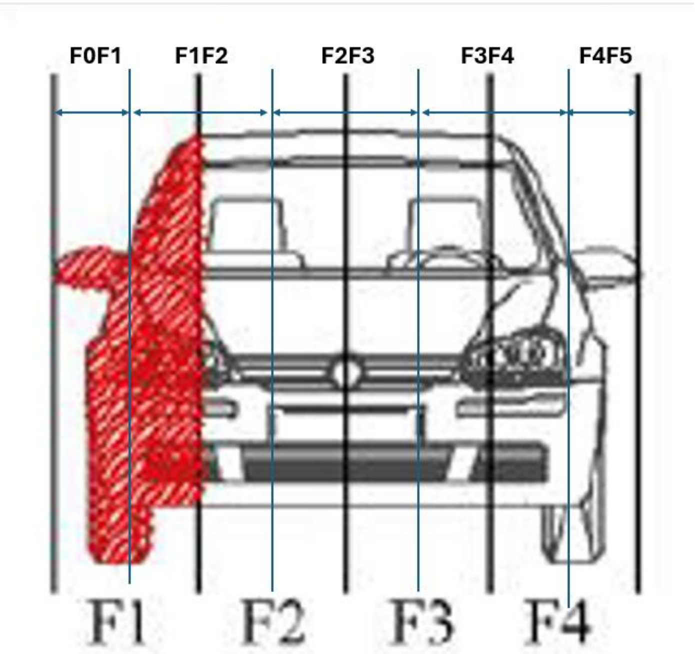


Figure 17: Deformation Zones (Federal Highway Research Institute (BASt) and Forschungsvereinigung Automobiltechnik e.V. (FAT), 2025)

The values F1, F2, F3 and F4 give values of depth in deformation, which is a one-dimensional value. That is why the first step will be to obtain the area of deformation on each deformation zone, estimating a continuous curve on each deformation zone, using these deformation zones and the total width of the car (W).

$$F0F1 = F1 * 0.125 * W$$

$$F1F2 = \left(F1 + \frac{F2 - F1}{2}\right) * 0.25 * W$$

$$F2F3 = \left(F2 + \frac{F3 - F2}{2}\right) * 0.25 * W$$

$$F3F4 = \left(F3 + \frac{F4 - F3}{2}\right) * 0.25 * W$$

$$F4F5 = F4 * 0.125 * W$$

Once the five manually calculated areas are obtained, the sum of these will result in the total deformation of the vehicle.

$$DEF = F0F1 + F1F2 + F2F3 + F3F4 + F4F5$$

This new variable does not offer satisfactory results for correlation, as an R-squared value of 0.12 shows a low correlation, which means that the hypothesis cannot be discarded. This result shows that some additional information is needed, and deformation by itself will not be enough to estimate the value for EES. However, the p-value, which is almost zero, shows that there is a real relation between both variables and, therefore, can be used in future calculations.

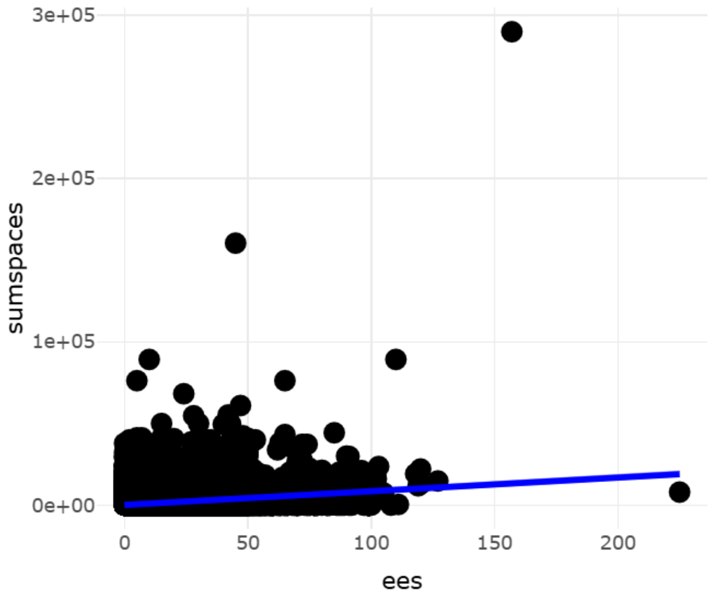


Figure 18: Deformation over EES graph.

5.2.2 POST-CRASH VELOCITY

The GIDAS database offers information about the vehicle speed in different moments of the crash, and also parameters related to the evolution of the velocity of the vehicle throughout the crash. However, the final velocity of the crash is not explicitly coded. Instead, the use of delta-v (one of the biggest correlating variables with ees in the database), the change of angle during collision (DWINK) and impulse angle (IMP) will be key for this final velocity, which is not expected to correlate, but can be used to obtain another crucial variable later on this thesis, which is the equivalent velocity of the change in kinetic energy. To obtain this value, it is important to understand which is the definition of the delta-V, and how it is calculated.

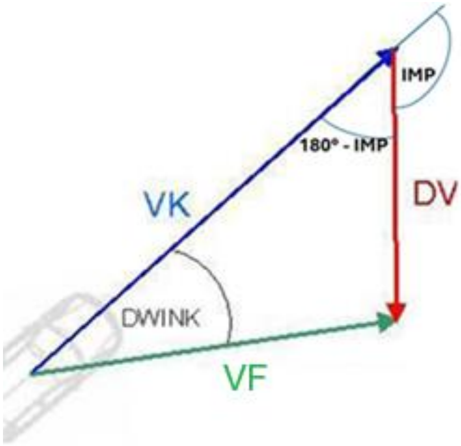


Figure 19: Definition of delta-V (Federal Highway Research Institute (BASt) and Forschungsvereinigung Automobiltechnik e.V. (FAT), 2025)

Once the values DV, DWINK and IMP are identified, final velocity can be obtained as shown in Equation 1.

$$\text{Equation 1: } VF = \sqrt{DV^2 + VK^2 - 2 * DV * VK * cos\,(180 - IMP)}$$

This variable returns the velocity of the car right after the collision. As expected, there is no correlation between this variable and EES, but it will be used for the next variable, which should be more relevant.

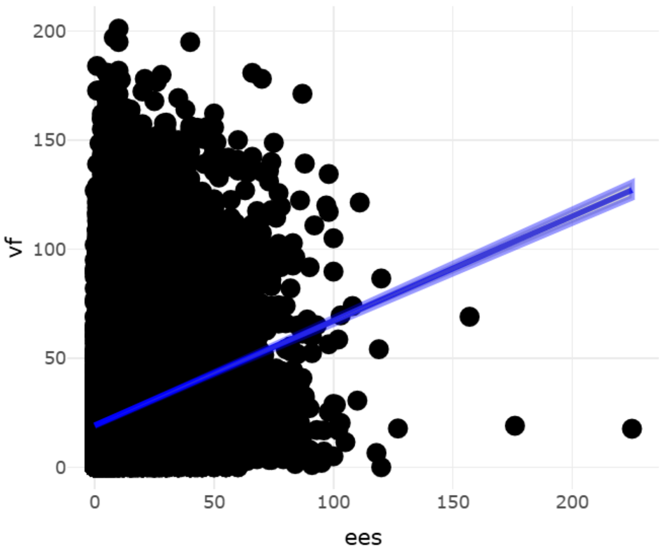


Figure 20: Final Velocity over EES Graph

5.2.3 EQUIVALENT VELOCITY

According to its definition, EES is the equivalent speed of kinetic energy in the deformation of the vehicle. This means that the value of EES should be possible to calculate as shown in Equation 2:

$$\textit{Equation 2: } K = \frac{1}{2} * m * EES^2$$

In a collision, the total amount of kinetic energy dissipated will be the total amount of the kinetic energy dissipated by each vehicle, which can be calculated as shown in Equation 3:

$$\textit{Equation 3: } K = \frac{1}{2} * \left(m_1 * (vk_1^2 - vf_1^2) + m_2 * (vk_2^2 - vf_2^2)\right)$$

The total energy of deformation of each vehicle depends on the relation of masses between the vehicles involved in the collision, suffering a higher deformation if the opposing car is heavier than the first car. That is why the energy distribution in the collision will be as shown in Equation 4:

$$\textit{Equation 4: } K_1 = K * \frac{m_2}{m_1 + m_2}$$

Using Equation 2, Equation 3 and Equation 4, the result of the equivalent velocity is:

$$U = \sqrt{\frac{\left(m_2 * \left(m_1 * (vk_1^2 - vf_1^2) + m_2 * (vk_2^2 - vf_2^2)\right)\right)}{m_1 * (m_1 + m_2)}}$$

This variable should, theoretically, be exactly the value of EES. However, there are other factors that influence the definition of the real EES variable, for example, the deformation of the car, which changes depending on the materials and the compatibility of the crash. Provided that the definition of EES is based on visual inspection, and most of the times, through photos of the crash, there is expected a variance between the calculation proposed and the real value.

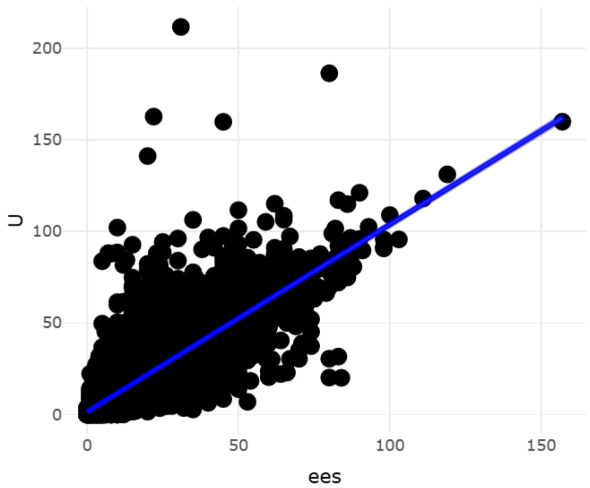


Figure 21: Equivalent Velocity over EES graph.

The correlation between these variables is more interesting than it looks in the graph. There is an estimation of a 1:1 relation between both variables. The R-squared value of 0.76 shows an important correlation and a p-value of ~0 will not allow to discard the hypothesis.

5.3 PREDICTIVE MODEL

Once the most relevant variables have been decided, the RandomForest library will allow using the variables chosen to generate a prediction of the EES value. The variables selected were the relative velocity (vrel), the age of the vehicle (bj_beginn), the deformation of the vehicle (sumspaces) and the equivalent velocity (U). Relative velocity was chosen because of its high correlation with EES, not being included in the calculation of the equivalent velocity (delta-v had a stronger correlation but that variable was included when calculating U). On the other hand the reason why the age of the vehicle was included is because, throughout the years, cars stiffness have been progressively changing, which supposes a different behavior when they suffer a car crash (Ellen L. Lee, 2014).

The table below shows the importance that the model gives to each variable during the decision-making process. It highlights the strong importance of the U value previously calculated and finds less important the model year of the vehicle.

	IncNodePurity
vrel	518238.07
sumspaces	349158.83
bj_beginn	43334.13
U	643818.27

Table 1: Importance of the Variables in the Predictive Model

The model will use all the cases which have a coded EES, and all the necessary variables defined as training values, which will also act as test values. This will be done this way for two reasons. The first reason is that, although it is possible to define criteria for identifying exemplary cases to be used as training data, thereby excluding those that may introduce noise or inaccuracies, the number of such qualifying cases is too limited to effectively train the model. The second reason is that the goal of the project is that EES values are coherent with each other, so, using this technique, it should be possible to identify outliers against the prediction of the model, note them, correct them and put them back in the database to perfect the training data.

This tool of the program enables the possibility of getting a prediction of all the cases, which is easier for the user, or each crash configuration/vehicle category separately, which delivers more precise and reliable results. In Figure 22, a really strong correlation can be seen between the EES and the predicted EES, which is close to the 1:1 line, marked in blue. However, a couple of outliers against the prediction of the model can be seen which allow to find cases that might be miscoded. By identifying those outliers, and correcting the EES values, a new iteration can be done finding different outliers against the prediction of the model up until all values become reliable.

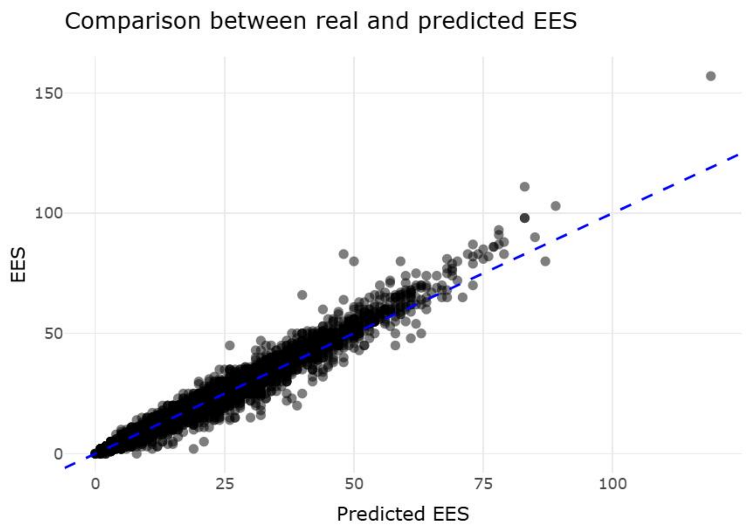


Figure 22: EES over Predicted EES graph.

The predicted values obtained in this prediction are not reliable yet because, as previously mentioned, the model is trained with both accurate and inaccurate data. That is why an additional tool must be developed to estimate the real EES value of the outliers against the prediction of the model found with this method. However, it is important to study the precision of the prediction, so it is possible to identify which parameters affect the model and how.

```
      Accuracy          Kappa  AccuracyLower  AccuracyUpper   AccuracyNull 
    0.24908094     0.22333170     0.24228890     0.25596724     0.05830377 
AccuracyPValue  McnemarPValue 
    0.00000000            NaN 
```

Table 2: Confusion Matrix Overall

For that reason, there are two tables that will be obtained. The first one is the overall of the prediction, where the most relevant statistical values are obtained to analyze the quality of the prediction and the process it followed. These values are obtained from the confusion matrix developed from the result of the prediction. The second one is the summary table of the cross validation where, through a 10-fold validation, parameters showing the accuracy of the prediction are shown.

```
Random Forest

15510 samples
    4 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 13958, 13959, 13960, 13958, 13959, 13959, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  2     4.291164  0.8393642  2.725880
  3     4.342306  0.8355073  2.753894
  4     4.392372  0.8316673  2.774922

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 2.
```

Table 3: Cross Validation Study

Finally, a box plot showing the error distribution will complement all the previous information with a visual and detailed analysis of the error of the prediction.

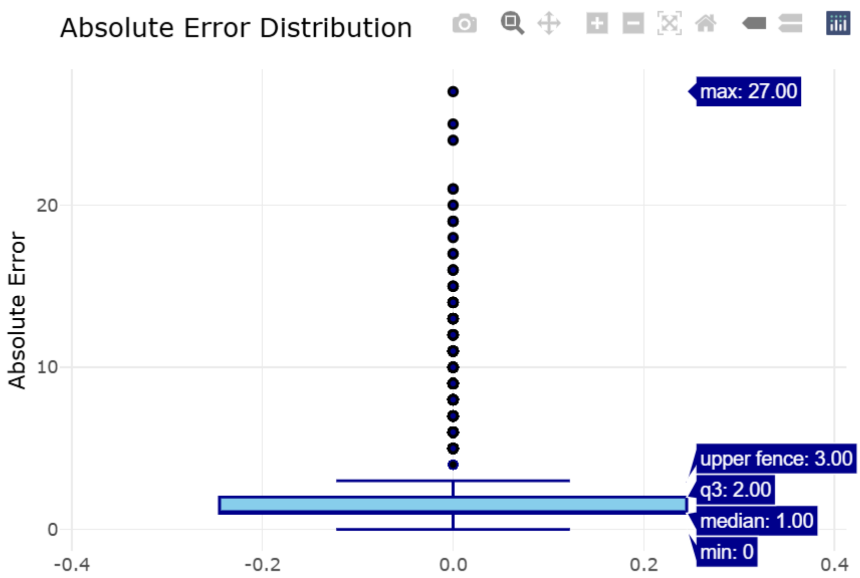


Figure 23: Box Graph of Absolute Error in Prediction

Chapter 6. CORRECTION OF THE DATABASE

Once miscoded cases have been identified, the EES Estimator tool is developed to create an estimation of the EES value this case should have. Then, these values are compared to the coded EES value in GIDAS and the predicted EES value obtained in the EES Analyzer tool. With that new comparison, conclusions can be obtained.

6.1 EES ESTIMATOR

EES Estimator is a function developed during this thesis that uses the EES-CNN tool of the program PC-Crash, with the help of Microsoft Copilot for the image recognition, in order to repeat the process of uploading the images of the cases to PC-Crash and retrieve the results several times without the need for a user who keeps doing this repetitive activity.

The tool is based on the following conceptual framework. The R program will take control of the mouse and the keyboard and perform a sequence of actions that react depending on the input the program receives. The sequence of the actions for this process to achieve the intended goal is explained below:

Preparation. Microsoft Copilot must be opened, and the rest of the programs closed, in exception of RStudio.

1. Open the Copilot window and open two selected images of each case to study with a command from RStudio. This can be done thanks to a function that receives the case number, the participant, and the position of the crash and returns the path for the two most relevant photos, supposedly showing the major part of the damage suffered during the crash. Copilot will validate whether the images can be used by the program or not using automated screenshots of the relevant information of the current screen.
2. Open PC-Crash with a command from RStudio, navigate through the menu to open EES-CNN tool and open the file explorer. In the file explorer the path previously

obtained will be introduced in the file bar to open the first image of those two, which is detected by the function as the most relevant.

3. After EES-CNN is done with its EES estimation, it returns a graph with a range of values where the EES should be for this crash, together with the corresponding confidence PC-Crash has on this range. The EES Estimator will then automatically generate screenshots of the resulting diagram. The result from PC-Crash will not generate any output that can be exported or copied, which means that the only way to receive data from the EES-CNN tool is to see it. That is why a screenshot is generated, and the next step will be to use the image recognition tool in Copilot, to get data that can be exported.
4. For the next step, the program closes PC-Crash, which leads back to Copilot, and pastes the screenshot with a prompt that tells the chatbot to learn the values shown in the screenshot. After that, it reopens PC-Crash and repeats the process getting the second most relevant image of the crash, which supposedly shows the most damaged area of the car but from a different perspective.
5. EES Estimator goes back to Copilot, after closing PC-Crash again, to paste this second screenshot and insert a different prompt that asks the AI to analyze the second graph, compare it to the first one, get the average of each bar and return only the range of values and the confidence with the format "XX, XX, XX", without any additional text.
6. The function will then copy Copilot's response and save it as a list of variables, which are saved as "min_eescnn", "max_eescnn" and "conf_eescnn" respectively and save them in the input data frame.

The reason why some validation is needed on the photos uploaded to PC-Crash is because in some cases the names of the images, which should define the case number, the participant and the angle of the photograph, are not defined correctly, showing some documentation, the road, the car from the inside or the car seen from the incorrect angle. There are other cases where the image will not exist by the name it should and other images which will pass the previous validation even though they do not show the most damaged part of the car, which will most likely lead to results with low confidence.

This process is repeated for as many cases as the user wants to estimate, and, each time the loop finishes a case, it updates an Excel table to save the calculated values in there, which allows the user to compare the values of coded EES in GIDAS, predicted EES in EES Analyzer and calculated EES in EES Estimator.

EES Estimator works as a separate program from EES Analyzer. However, EES Analyzer has the option to export up to 1000 rows containing the data with the biggest difference between coded EES and predicted EES. These 1000 rows could be more, but this number of repetitions was chosen as a maximum because, taking the program approximately 2 minutes to complete one estimation, it is expected to complete 1000 estimations in less than 34 hours. More cases would be too much time, and it is not the main goal of this study.

Chapter 7. RESULTS

After developing both EES Analyzer and EES Estimator, now all the tools required for this project are available, therefore the analysis of the GIDAS database can be conducted. This analysis will include all the goals proposed in section 4.2.

7.1 CONTRIBUTING FACTORS TO EES CODING ERRORS

An analysis was conducted using approximately 20,000 cases from the GIDAS database, focusing exclusively on frontal collisions involving only two vehicles and excluding multi-impact events. To estimate the error in EES codings, a Random Forest model (5.3. Predictive Model) was trained on the same dataset, with the error defined as the difference between the predicted and the coded EES values.

The results reveal several important insights. First, the moderate linear correlation between the EES value and its associated prediction error ($R^2 = 0.3$) suggests that higher EES values are more prone to larger deviations. This finding raises concerns about the robustness of EES codings in high-velocity crashes. The increasing error with EES magnitude may reflect limitations in the model's ability to capture complex crash dynamics, or inconsistencies in human coding practices under more severe conditions.

There are two reasons why this might be happening. The first reason is that there is a relatively small number of high-speed crashes in GIDAS, which turns into a lack of information for the model to learn from those cases. That means that the model will learn from the lower speed crashes and, therefore, will not consider some variables which might behave differently when the speed is higher. The second reason is that when crash speed exceeds the typical crash test speed, the EES for this crash become more difficult to code. This happens because EES values are usually based on crash test results.

Second, the variation in residual errors across vehicle types, ranging from 2.395 in super compacts to 4.437 in compacts and 4.343 in SUVs, indicates that structural and design differences among vehicles may not be fully accounted for in the current coding approach. These discrepancies suggest that the vehicle category plays a significant role in the accuracy of EES predictions and should be considered in future model refinements. Moreover, they highlight the need for more clustered crash studies depending on the vehicle classification.

In addition, an unusually strong correlation was found between EES and Delta-V. While a degree of correlation is expected due to the physical relationship between impact severity and vehicle velocity change, the strength of this linear relationship raises conceptual concerns. EES is intended to represent the energy dissipated in a crash, reflecting the physical damage of the vehicle, and it is not intended to rely on a single variable, as it looks in this case. The observed correlation suggests that, in practice, Delta-V may be influencing more than it should on the coding of EES.

This finding aligns with critiques such as those found in the Breitlauch model (Breitlauch, Junge, Erbsmehl, Sandner, & van Ratingen, 2023), which emphasize the importance of maintaining the conceptual distinction between crash reconstruction variables and outcome-based metrics like EES. The purpose of EES is not to replicate the crash dynamics but to quantify the consequences. Therefore, while predictive tools can assist in estimating EES values, they must be used with caution to avoid replacing expert judgment.

Ultimately, this study underscores the importance of critically evaluating the tools and processes used in crash data analysis. As predictive modeling becomes increasingly integrated into safety research and operational workflows, maintaining the balance between efficiency and conceptual rigor will be essential to ensure that data-driven insights remain valid, reliable, and meaningful.

7.2 INFLUENCE OF CRASH CONFIGURATIONS ON ACCURACY OF EES PREDICTIONS.

When generating a prediction of the EES value using a model generated with the RandomForest package, the crash configuration is not defined as a variable to be taken into account to generate the prediction. However, this parameter is key to understanding how the car is affected after a car crash. To check how relevant this factor is in the prediction model, a prediction was made using all the cars which have a coded crash configuration value according to the crash configurations specified by Brumbelow (Brumbelow, 2019). Then, a prediction was made for every single crash configuration separately, to check how the importance of the reference variables changes and how the error and accuracy of the model evolves.

7.2.1 DATASET OVERVIEW AND MODEL PERFORMANCE

The dataset analyzed in this study consists of a subset of the more than 20.000 cases used for the complete project, using only the cases with a valid Brumbelow-coding. Therefore, it will be used a dataset of 2288 two vehicle frontal crashes. The distribution of those cases is shown in Table 4:

	Samples
Generic Case	2288
Large Overlap	467
Moderate Overlap	675
Small Overlap	394
Perpendicular	254
Oblique Center	144
Oblique Corner	317

Table 4: Number of Cases per Configuration

As explained in 5.3. Predictive Model, the variables used to generate this estimation are the relative velocity, the year when the car was made, the deformation area in the impact and the equivalent velocity of the kinetic energy calculated in 5.2.3. Equivalent Velocity.

The table below shows the importance that the predictive model gave each variable when generating the decision trees. This number will be higher the more data is included in the decision tree, so there is no direct conclusion with that information. However, the percentual distribution of the four variables will give a more meaningful understanding of how the model treats each variable for each crash configuration.

	Importance vrel	Importance year	Importance deformation	Importance U
Generic Case	130166.65	75003.33	25460.52	139030.23
Large Overlap	25693.116	5735.215	17630.811	29584.422
Moderate Overlap	47478.36	14021.14	32989.54	54770.22
Small Overlap	23380.98	3616.86	11997.78	24272
Perpendicular	4842.529	1578.907	3075.107	6251.961
Oblique Center	4497.425	1145.935	2489.005	4838.245
Oblique Corner	11134.166	3488.596	6603.907	11940.058

Table 5: Importance of Predictors for each Crash Configuration

To evaluate the predictive performance of the machine learning models, three regression metrics were studied: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R-Squared). The table below shows how the predicted EES value and the one coded in GIDAS correlate with each other, according to these metrics.

	MAE	RMSE	R-Squared
Generic Case	2.951053	4.721384	0.8743488
Large Overlap	2.470125	4.20695	0.9092729
Moderate Overlap	3.088576	5.327708	0.8873415
Small Overlap	3.434385	5.291476	0.8463373
Perpendicular	2.908069	3.871057	0.7929549
Oblique Center	3.559649	5.307713	0.7492057
Oblique Corner	3.195927	4.303282	0.8481095

Table 6: Model Performance Metrics for each Crash Configuration

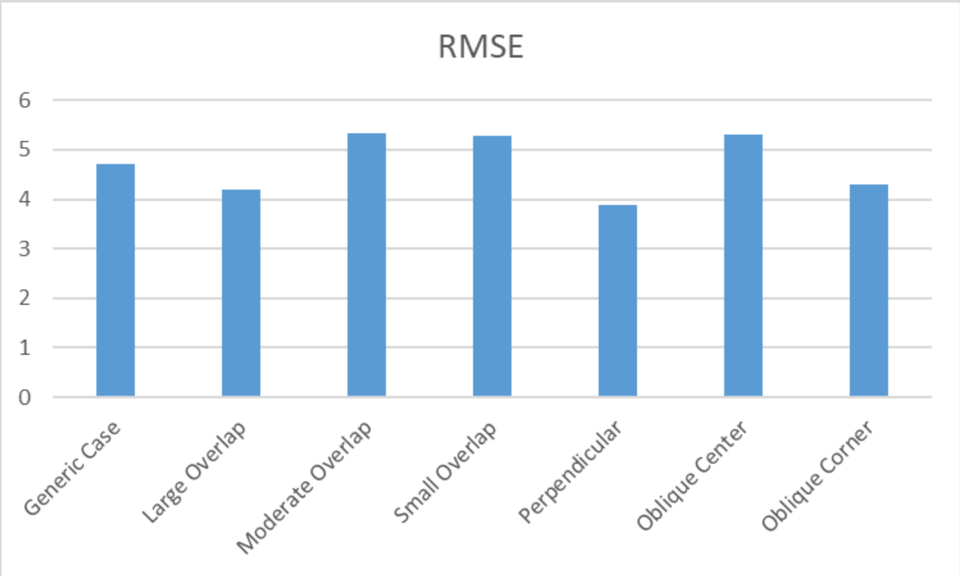


Figure 24: Bar Chart Comparing RMSE by Configuration

This comparative analysis revealed that the generic case performed better than expected. The model generated with the generic case had an intermediate behavior when compared to the rest of the configurations. That could mean that some configurations have better behavior when predicted with the generic case but also could mean that there is a high presence of outliers against the prediction of the model in these configurations, meaning that it could be another contributing factor for EES coding error.

It is also concerning the unexpectedly high value of moderate overlap models. As EES is typically based on crash test results, which are most of the time large and moderate overlap, these models are expected to have the best performance. However, even though the large overlap does behave as expected, moderate overlap has a value which is higher than it should be and, therefore, more studies will have to be conducted with cases with moderate overlap.

7.2.2 Prediction Accuracy and Error Visualization

To compliment the numerical performance metrics, scatter plots will show the relationship between predicted and actual EES values.

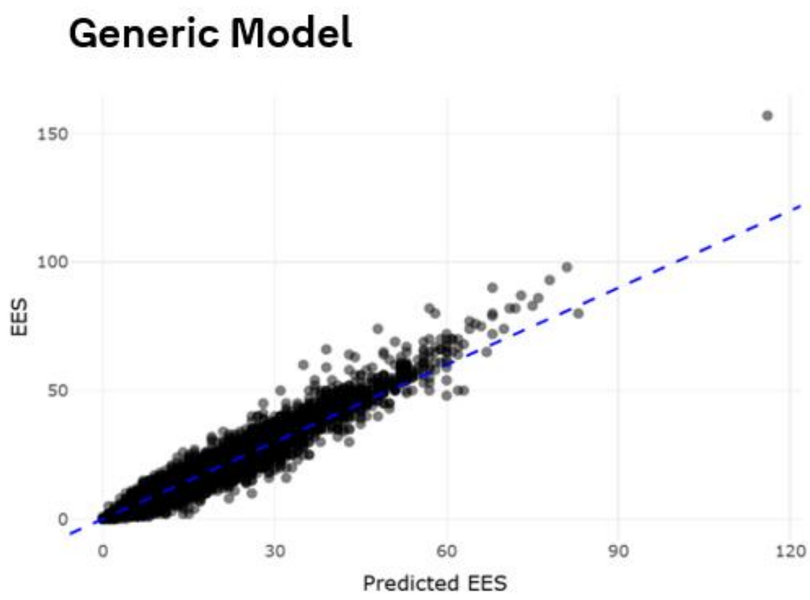


Figure 25: Comparison of Predicted EES vs. GIDAS - Generic Model

With the scatter plot graphs, further conclusions can be made, such as whether the configurations that performed worse were because of the model or the presence of outliers, when compared to its respective model. Oblique center configuration showing clear outliers against the prediction of the model means that correcting the miscoded EES cases could improve the performance of the model. However, small and moderate overlap showing disperse clouds of points for higher EES values means that the reason for high error could be for both not effective model and presence of miscoded cases, as neither of those hypotheses could be discarded. This result is highly unexpected. The methodology of EES coding consists of comparing the crashed vehicle to comparable crash tests, which means that the models for large overlap and moderate overlap should be the most consistent models generated, because those are the configurations that are most present in crash tests, and the rest of them are expected to be worse than these two. However, moderate overlap showing a disperse cloud with a relatively high error is unexpected and will need further study in the future.

Even though Figure 24 showed that the generic model performed better than expected, as explained above, its scatter plot (Figure 25) shows a cloud of points which is wider than all of the other scatter plots for particular configurations. This proves that, in order to define the EES value of a car crash, it is necessary to compare the crash with crashes with the same crash configuration. The scatter plot for each of the crash configurations studied are shown in ANEX II.

To further assess prediction consistency, the distribution of absolute prediction errors is examined using boxplots. These visualizations display the range, interquartile spread, and median error values for each configuration, as well as the generic case.

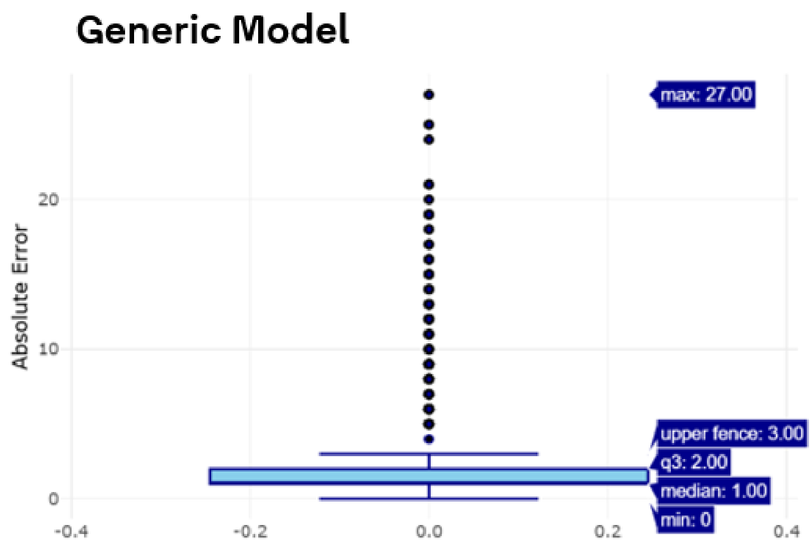


Figure 26: Error Distribution - Generic Model

These error plots show the interquartile range (IQR) of the absolute error between coded EES and predicted EES, which is shown as a box and includes 50% of the cases. The whiskers show the rest of the data which is not considered an outlier (following the rule of 1.5 x IQR), and the rest of the points that appear in the graph are those considered outliers. This way, the amount of points and the value of the upper fence can give clues about the error in the model, where a high upper fence means a high error for valid cases, and the error in the coding, where several points over the upper fence should mean a high number of outliers. The box plots for each of the crash configurations studied are shown in ANEX III.

From these graphs, three different conclusions can be obtained:

- The upper fence of small overlap is set as 3, which means that most of the deviations from the original coded values were outliers against the prediction of the model, with values higher than 1.5 x IQR (interquartile range), which can be seen in its box plot (Figure 42).
- The higher fence for moderate overlap is set as 5, which shows a generic dispersion in the model, where most of the cases are included inside the box or in the whiskers, and only a few as outliers compared to the prediction of the model. That means that the model generated using this crash configuration will not be reliable yet and, therefore, a more in-depth study will be needed.
- There is a higher number of outliers shown in the generic case box plot (Figure 26), which is expected, however the value of those outliers, when compared to their respective value, is noticeably higher than in the rest of the cases, with many values higher than 15, which is the maximum of all of the rest of the box plots and appears in Figure 50.

Another relevant aspect to consider is that, with each prediction, certain values may remain unpredicted due to the program's limitations in making specific determinations. That is the reason there is a value in the generic case graph which has a coded EES of more than 150 kph, which does not appear in any other graph. This means that when predicted for each configuration separately, less values are obtained, therefore the generic prediction will always be needed to be able to obtain more values.

7.2.3 VARIABLE IMPORTANCE

To understand the difference between all the crash configurations, it is necessary to understand how the model generates the prediction, in this case, by studying how important the program understands each variable is and, therefore, which value the program assigns to this variable in order to determine the predicted value. Graphs showing the distribution of the importance for each model can be seen below:

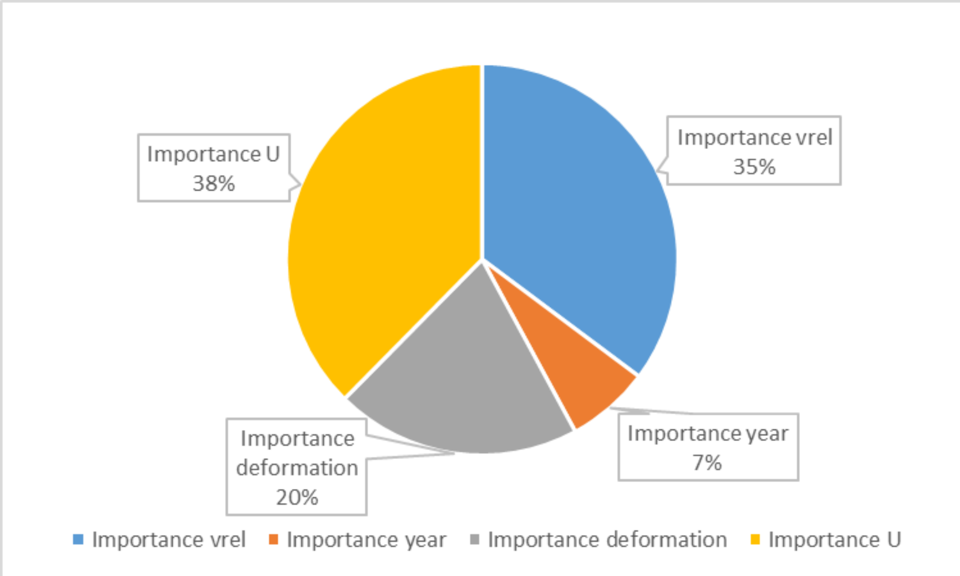


Figure 27: Importance Distribution - Generic Model

When RandomForest generates a model, it assigns values to the importance that the model gives to each variable. These values depend on the amount of data that the model has for training. For that reason, the importance of the variables only gives interesting information about the model when seen as a percentage of the total. The pie charts for each of the crash configurations studied are shown in ANEX IV.

When the deformation variable was first introduced into the model, it was expected to play a significant role, possibly even rivaling relative velocity as one of the most influential predictors of EES. However, the results show that its importance is lower than anticipated. While it does rank above vehicle age (which was expected to have minimal impact), it does not come close to the relevance of relative velocity or equivalent velocity (U).

That said, deformation still adds meaningful value to the model. Unlike the other key variables, which focus on the dynamics of the crash, deformation brings in a structural perspective. It reflects how the vehicle physically responded to the impact, something that neither velocity nor age can capture directly. In fact, aside from vehicle mass, present in the calculation of equivalent velocity, the model lacks any other input that speaks to the vehicle's structural characteristics. So even if deformation is not the top predictor, it helps round out the model by offering a different kind of insight, one that enhances precision and supports a more complete understanding of crash severity.

On the other hand, the equivalent velocity stands out as the most consistently important variable throughout the entire study. Its relative importance remains remarkably stable, ranging from 36% in oblique corner impacts to 40% in perpendicular collisions. In every scenario analyzed, it ranks as the most influential factor in the model. This consistency suggests that equivalent velocity plays a leading role in predicting EES values, regardless of crash configuration, and reinforces its reliability as a key input in the modeling process.

7.3 CORRECTION OF MISCODED VALUES

The process of correction of the database is based on three phases: filtering and prediction using EES Analyzer, detection of potential outliers and first visual inspection and running PC-Crash, manually or using EES Estimator.

Section 7.2 showed that there are many outliers against the prediction of the model in oblique center configuration, so those outliers need a closer look. This objective is achieved through the EES Analyzer tool, which enables straightforward identification. After defining the filters, a prediction is executed:

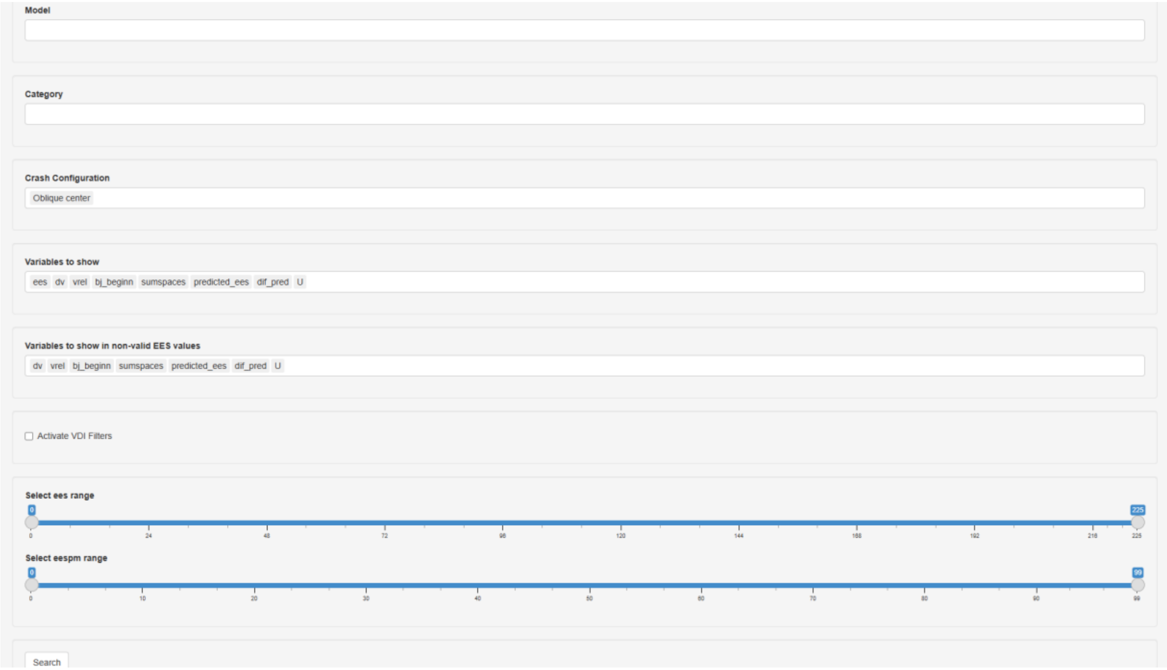


Figure 28: Filters for EES Correction for Oblique Center

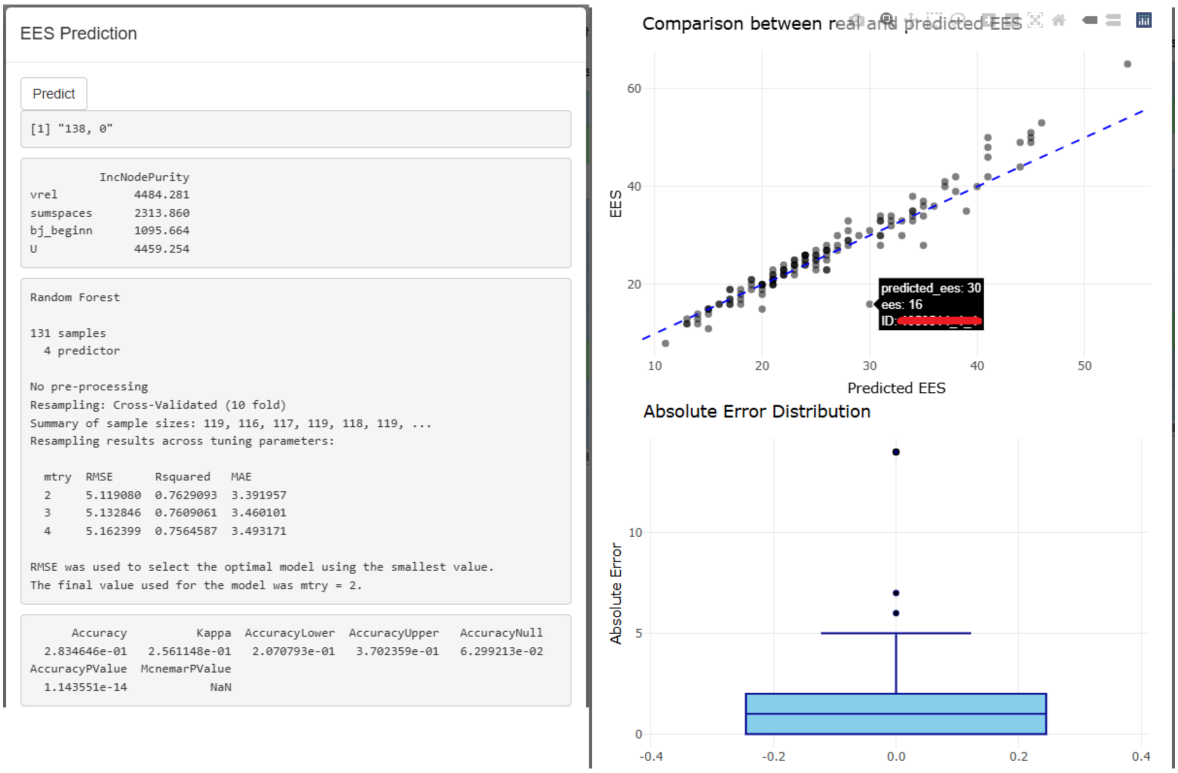


Figure 29: Outlier Identification from EES Analyzer

Although the identification process can be applied to multiple cases individually, this study focuses on a single instance in which the predicted EES value is 30, whereas the corresponding coded value in the GIDAS database is 16.

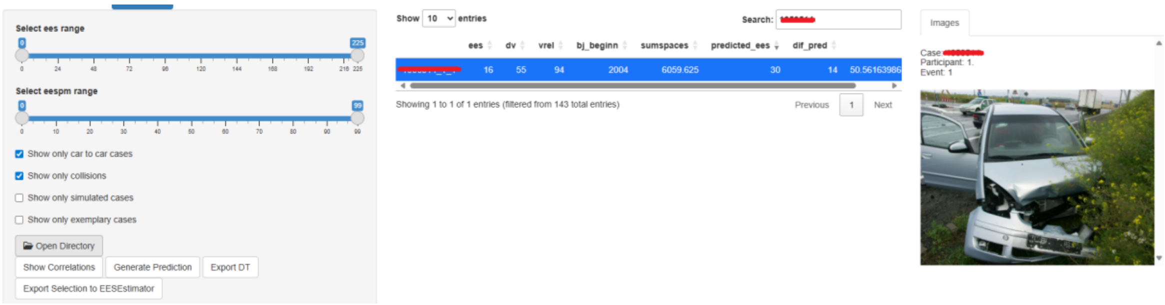


Figure 30: Potential Outlier Identified in the Database

At this point, there are two ways to apply the EES-CNN function of PC-Crash to this case. The first one is to do it manually, clicking on Open Directory and sending manually the relevant images to PC-Crash. This way is more reliable than the second one, but drastically more time-consuming than the second option, which consists of clicking Export Selection to EES Estimator. This option is better when the study consists of more cases, as it will work automatically. As in this case, there is only one car that needs to be studied, the estimation will be done manually.

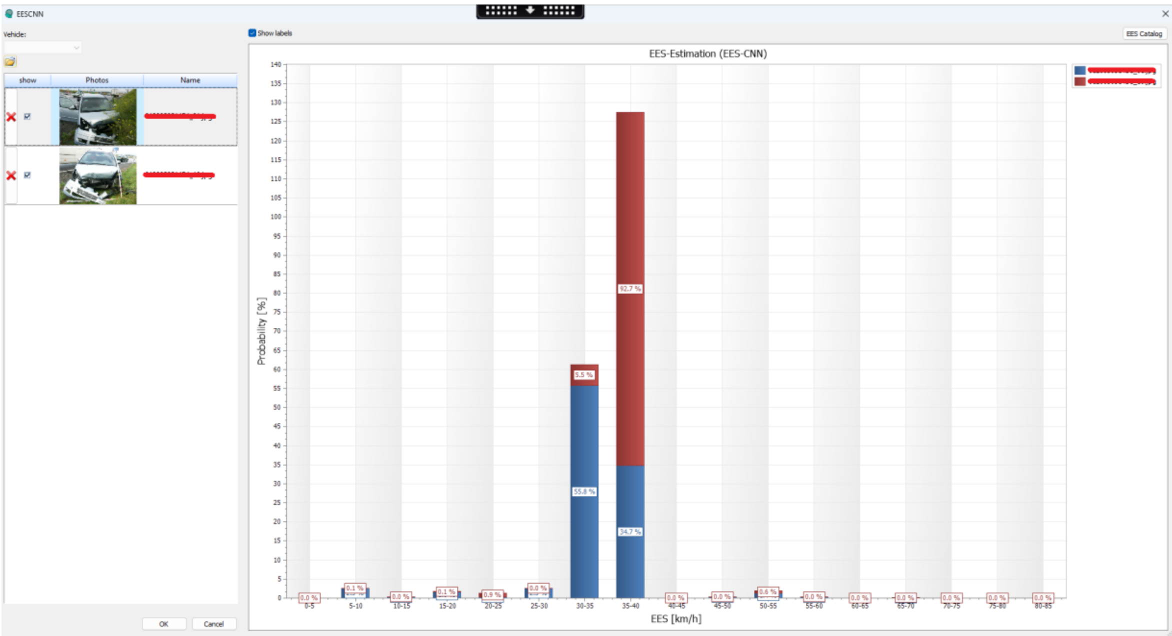


Figure 31: PC-Crash Visual Estimator Output

Results show that the coded EES value for this case is 16 kph, but the prediction model has estimated a value of 30 kph. So, this case was defined as a potential outlier and sent it to PC-Crash. This software estimated a value of 35-40 kph with a confidence value of 63.7%, and a value of 30-35 kph with a confidence value of 30.65%. These two percentages were calculated as an average of the estimation of each one of the two images, and the results it shows are a proof that this case might be a miscoded case, and its real value should be closer to 35 kph than to the 16 kph that were originally coded.

7.4 COMPARISON BETWEEN GIDAS, EES ANALYZER PREDICTION AND EES ESTIMATOR RESULTS

To evaluate the relative accuracy of three different approaches for estimating the Equivalent Energy Speed (EES), a comparative analysis was conducted involving:

- GIDAS EES coded value
- Predicted EES from EES Analyzer
- Estimated Range from EES Estimator, defined by intervals of 5 km/h.

For the comparison, only estimated ranges with a confidence higher than 70% are considered.

7.4.1 VISUAL COMPARISON

The first visualization (Figure 32) presents a scatter plot where the X-axis represents the EES value (in km/h) and the Y-axis corresponds to the case index. Each case includes three elements: the GIDAS-coded EES, the predicted EES, and a horizontal bar indicating the range estimated by EES-CNN. This plot enables a direct visual comparison of the three values for each case, allowing a direct visual comparison between the values. When comparing only two EES variables, it is not possible to identify which one is more likely to be wrong. Therefore, a comparison between the GIDAS EES, the predicted variable using EES Analyzer and the estimated range using EES Estimator must be compared to each other in order to identify whether this project is increasing the accuracy of EES or not.

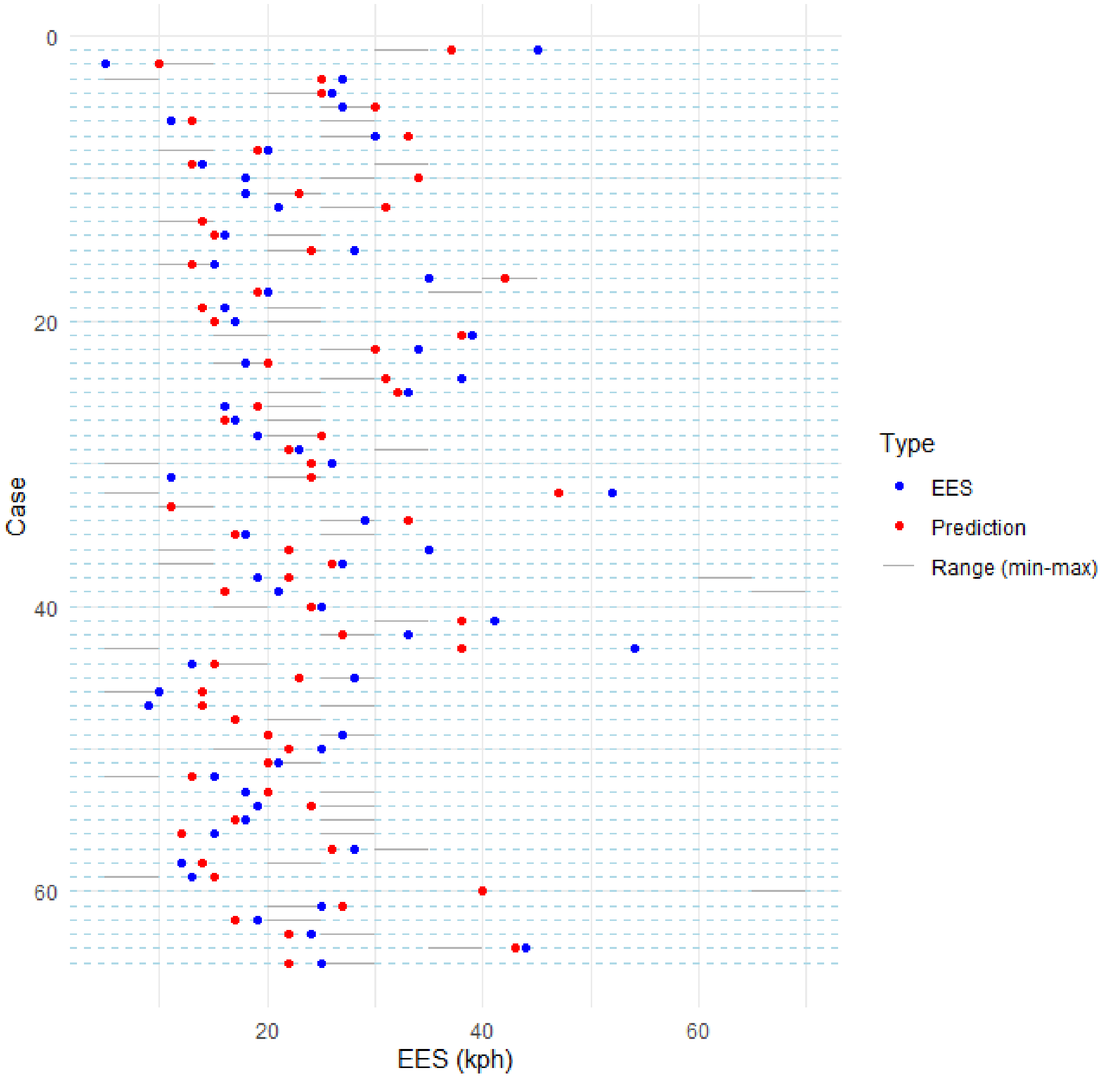


Figure 32: Dumbbell Plot: Real Values vs Estimated Range

The graph shows how the value of predicted EES has a higher tendency to get closer to the estimated range than GIDAS EES, which means two things. First, predicted EES can be a strong tool to identify potential outliers, as the cases where the difference between both EES values is higher also tend to be those where GIDAS EES is further from the estimated range. Secondly, predicted EES and estimated range tending to similar values implies that the prediction tool is reliable in order to determine EES values.

7.4.2 DEVIATION ANALYSIS

To quantify the deviation of each method from the estimated range, a heatmap was generated (Figure 33). Each cell in the heatmap represents the absolute difference between the center of the estimated range and either the coded or predicted EES, which is helpful to understand which EES value is closer to the center of the estimated range of PC-Crash. The heatmap is organized into two columns: one showing the deviation of the coded EES from the center of the range defined by PC-Crash, and the other showing the deviation of the predicted EES from that same center of the range. This allows for a systematic comparison of both methods in terms of their alignment with the estimated range.

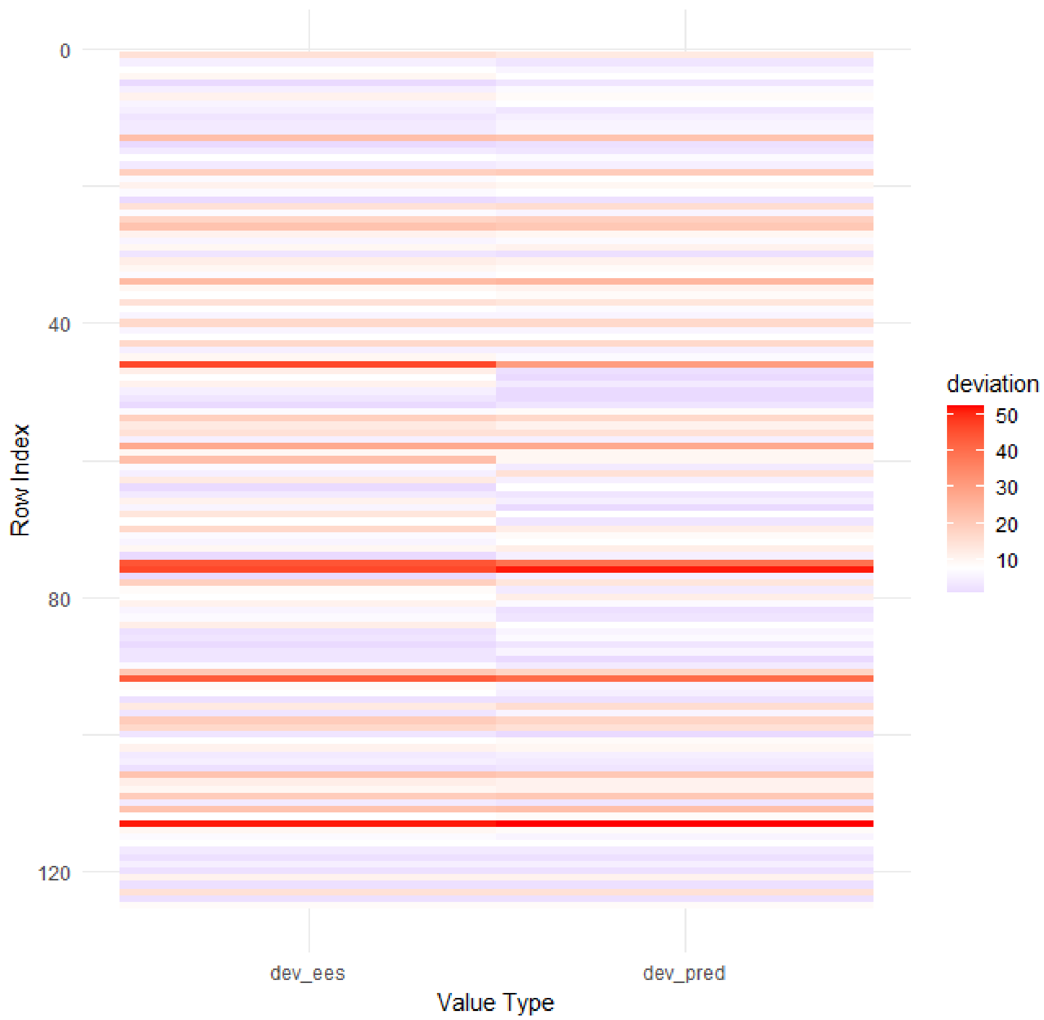


Figure 33: Heatmap of Deviations from Estimated Range Center

The plot shows really similar behavior in the deviation from the center of the range estimated by PC-Crash. However, the deviation from this center of the range is slightly lower, and therefore, slightly better, in the predicted values using RandomForest than in those coded in GIDAS.

This study sets out to explore how predictive modeling could support and improve the coding of EES values in crash data, using a Random Forest model trained on a large subset of GIDAS cases. The results show a promising direction: in most cases, the predicted EES values and the estimated EES ranges are more consistent with each other than with the originally coded EES values. This pattern, visible in Figure 33, suggests that the model is not only capable of producing reasonable estimates but may also help identify when a coded value falls outside expected bounds.

One of the key strengths of this approach lies in the use of the estimated range. Rather than relying solely on a single predicted value, the estimated range provides a context within which both the coded and predicted values can be evaluated. When the coded EES falls far outside this range, but the predicted value remains within it, it raises a flag: perhaps the prediction is closer to reality, and the coded value deserves a second look. This kind of insight could be especially useful in quality control or in cases where manual coding is uncertain or inconsistent.

However, it is important to recognize the limits of automation. EES is not just a number derived from crash mechanics, but it is meant to reflect the physical consequences. As stated by Breitlauch et al., (2023), EES should remain conceptually distinct from variables like Delta-V, which describe the dynamics of the crash itself. If predictive tools begin to dominate the coding process without human oversight, there's a risk of losing that distinction.

So, while the Random Forest model and the estimated range offer valuable support, they should be seen as tools to assist, not replace, expert judgment. Their role is to guide coders, highlight inconsistencies, and improve efficiency. In fact, the integration of predictive modeling and range-based validation could lead to a more consistent and transparent coding process, especially in large datasets where manual review is time-consuming.

Looking ahead, these findings open the door to developing smarter tools within platforms like the EES Analyzer. With careful design, such tools could help coders make better decisions, reduce errors, and ensure that EES values remain meaningful and reliable. But any future implementation must respect the original purpose of EES and maintain the balance between technological support and human expertise.

Chapter 8. CONCLUSIONS AND FUTURE PROJECTS

This chapter of the report serves as a summary of all the goals achieved and the conclusions taken from them, along with future steps to take in order to achieve the ultimate goal of improving the accuracy of EES coding in order to improve the reconstruction of car crashes.

8.1 CONCLUSIONS

One of the core components of this thesis was the development of a software application using Shiny (R), designed to facilitate interaction with the GIDAS (German In-Depth Accident Study) database. The motivation behind this development was to streamline the process of querying, visualizing, and analyzing traffic accident cases, which are often complex and data-rich.

The tool provides an intuitive interface that allows users to filter accident cases based on specific characteristics, such as vehicle type, collision configuration, or injury severity. This filtering capability enables researchers and analysts to focus on subsets of data that are most relevant to their investigations, significantly reducing the time and effort required to manually sift through large datasets.

In addition to filtering, the software includes a search function that allows users to locate specific cases directly. Upon selecting a case, the tool displays a preliminary image associated with the accident, offering immediate visual context. Furthermore, it provides direct access to the case directory, where users can explore additional images and documentation related to the incident.

To support statistical analysis, the application incorporates a linear regression module, which can be applied to any two variables within the filtered dataset. This feature enables users to quickly identify potential correlations and trends without needing to export data to external tools.

A particularly innovative aspect of the software is its ability to predict Equivalent Energy Speed (EES) using a random forest model. This predictive functionality adds a layer of analytical depth, allowing users to estimate EES values based on other case parameters, which can be especially useful in scenarios where direct measurements are unavailable.

Finally, the tool includes data export capabilities, allowing users to either export the entire filtered table or selected cases to Excel format. This functionality supports seamless integration with a second software tool developed during the thesis, named EES Estimator, ensuring a smooth workflow between data exploration and detailed analysis.

The second major contribution of this thesis was the development of EES Estimator, a software tool designed to automate the interaction with the EES-CNN module of PC-Crash. This tool was created to address a significant bottleneck in the workflow: the manual processing of large volumes of accident cases to obtain Equivalent Energy Speed (EES) predictions.

EES Estimator streamlines this process by sending a large batch of cases to the EES-CNN tool, validating the associated images, uploading them, and retrieving the predicted EES values. This automation was made possible through the use of R programming, and with the support of Microsoft's large language model: Copilot, which facilitated the integration and handling of the various steps involved in the communication with the external tool.

The manual execution of this process for hundreds of cases would be extremely time-consuming and error-prone, making it impractical for large-scale studies. By automating the workflow, EES Estimator significantly reduces the time required to obtain results, while ensuring consistency and reproducibility across cases.

This tool plays a crucial role in enabling high-throughput analysis of traffic crash scenarios, allowing researchers to focus on interpretation and modeling rather than data preparation and manual input. It complements the previously developed EES Analyzer, forming a robust pipeline for accident data exploration, prediction, and export.

To evaluate the accuracy of EES codings and detect the model's limitations, an experiment using 20,000 frontal two-vehicle collisions from the GIDAS database revealed key limitations in EES codings. A Random Forest model showed that prediction errors increase with EES magnitude, suggesting reduced reliability in high-speed crashes. Differences in error across vehicle types indicate that structural variations are not fully captured, pointing to the need for vehicle-specific modeling.

Additionally, a strong correlation between EES and Delta-V raises concerns about the conceptual integrity of EES, which should reflect crash consequences rather than be driven by a single dynamic variable. These findings support existing critiques and emphasize the importance of maintaining expert judgment and conceptual clarity when using predictive tools in crash analysis.

Another part of the study explored how different crash configurations affect the accuracy of EES predictions using the EES Analyzer. Surprisingly, the generic model—trained on all configurations—performed better than expected, showing intermediate behavior compared to specific configurations. This suggests that some configurations may be better predicted using a generic model, or that certain configurations contain more outliers against the prediction of the model, contributing to EES coding errors.

Scatter plots revealed that configurations like oblique center contain clear outliers against the prediction of the model, indicating that correcting miscoded cases could improve model performance. In contrast, small and moderate overlap configurations showed dispersed point clouds at higher EES values, suggesting that both model limitations and coding inconsistencies may be responsible for the errors. Notably, moderate overlap, expected to be among the most reliable due to crash test data, showed unexpected dispersion and will require further investigation.

Boxplots of absolute prediction errors provided additional insights. Small overlap had most deviations classified as outliers, when compared to its respective predicted value, while moderate overlap showed a broader error distribution, indicating model unreliability. The generic model had the highest number and magnitude of outliers from the prediction made using this model, reinforcing the need to match crash configuration when estimating EES.

Another important observation was that some high EES values were only predicted in the generic model, due to limitations in configuration-specific models. This highlights the necessity of using the generic model to ensure broader coverage.

Finally, analysis of variable importance showed that equivalent velocity was the most influential predictor across all configurations, with consistent relevance. Deformation, while less important than expected, still contributed valuable structural insight, complementing dynamic variables and enhancing the model's interpretability.

The next phase of the project consisted of the correction of potentially miscoded EES values was structured into three phases: filtering and prediction using EES Analyzer, identification of outliers against the prediction of the model through visual inspection, and validation using PC-Crash, either manually or via EES Estimator. One specific case was selected for closer examination, where the predicted EES was 30 kph, while the coded value in the GIDAS database was only 16 kph.

To validate this discrepancy, the case was processed manually through PC-Crash. The software estimated an EES of 35–40 kph with a confidence of 63.7%, and 30–35 kph with 30.65% confidence, based on two image inputs. These results strongly suggest that the original coding of 16 kph may be inaccurate, and that the true EES value is likely closer to 35 kph.

This example demonstrates the effectiveness of combining automated prediction tools with expert validation to identify and correct inconsistencies in crash data, ultimately improving the reliability of EES coding.

Lastly, to assess the relative accuracy of different approaches for estimating Equivalent Energy Speed (EES), a comparative analysis was conducted using three sources: the coded EES values from the GIDAS database, the predicted values from EES Analyzer, and the estimated ranges from EES Estimator (based on PC-Crash), considering only estimates with over 70% confidence.

A visual comparison showed that, in many cases, the predicted EES values aligned more closely with the estimated ranges than the original coded values. This suggests that the Random Forest model may offer more reliable estimates and could help identify potentially miscoded cases. A heatmap analysis confirmed this trend, showing slightly lower deviations between predicted values and the center of the estimated range compared to the coded values.

The use of estimated ranges adds valuable context, allowing both coded and predicted values to be evaluated against a confidence-based benchmark. When a coded value falls outside this range but the prediction remains within it, it signals a possible inconsistency worth reviewing.

However, while predictive tools enhance efficiency and consistency, they must not replace expert judgment. EES is meant to reflect crash consequences, not just dynamics, and must remain conceptually distinct from variables like Delta-V. As such, predictive models should be seen as supportive tools that guide coders and improve data quality, not as substitutes for human expertise.

These findings point toward the potential for smarter, more reliable coding workflows, especially in large datasets, while emphasizing the need to preserve the original intent and integrity of EES as a safety metric.

8.2 FUTURE PROJECTS

Throughout this thesis, tools have been developed, and analyses have been carried out which aimed at improving how we work with crash data, particularly in estimating and validating Equivalent Energy Speed (EES). While the results have been promising, they also open the door to several future developments that could make these tools even more useful and reliable.

One of the most immediate next steps is to refine the predictive models used in EES Analyzer and EES Estimator. As seen in the analysis, certain crash configurations, like moderate overlap, did not behave as expected. This suggests that the models could benefit from more detailed study of each classification of crash types, considering a broader set of input variables, especially those that reflect the structural characteristics of vehicles.

In addition, varying the predictors revealed that, when removing the deformation from the predictors, the accuracy of the prediction was increased. Although that would support the point made in this thesis that there might be too much reliance in dynamics in the EES codings in GIDAS, no further study was made to support that supposition.

Another area for improvement is the integration between automated tools and manual validation processes. While automation saves time, especially when working with large datasets, expert judgment remains essential. A hybrid approach, where predictive tools assist but do not replace human review, could strike the right balance between efficiency and accuracy.

There is also potential to develop quality control features within EES Analyzer. These could automatically flag cases with unusually high prediction errors or inconsistencies between coded and predicted values, helping analysts focus their attention where it is most needed.

The use of estimated ranges, rather than single predicted values, has proven to be a valuable addition. It provides context and helps identify when a coded value might fall outside expected bounds. Expanding this concept, perhaps by incorporating confidence intervals or probabilistic outputs, could make the predictions even more informative and trustworthy.

Finally, and perhaps most importantly, future work must continue to respect the conceptual integrity of EES. As predictive modeling becomes more common, we must ensure that EES remains a measure of crash consequences, not just a reflection of crash dynamics. This means keeping a clear distinction between variables like Delta-V and outcome-based metrics and always using technology to support the expertise of crash analysts.

In short, the tools and methods developed here are just the beginning. With further refinement, they have the potential to make crash data analysis more consistent, transparent, and scalable. But in the future, it will be crucial to maintain a thoughtful balance between innovation and the human insight that gives meaning to the data.

Chapter 9. BIBLIOGRAPHY

Breitlauch, P., Junge, M., Erbsmehl, C., Sandner, V., & van Ratingen, M. (2023). Converting Post-crash Deformation into Energy Equivalent Speed (EES) Toward an Objective Measure. *Journal of Automobile Engineering*.

Brumbelow, M. L. (2019). Front Crash Injury Risks for Restrained Drivers in Good-rated Vehicles by Age, Impact Configuration, and EDR-based Delta V. *IRCOBI conference.*

Campbell, K. (1974). Energy Basis for Collision Severity.

Ellen L. Lee, P. J. (2014). *Increase in Front, Rear and Side Stiffness Coefficients in the Past Twenty Years Necessitates New Representative Database.*

Eric R, T., & Nolan, J. M. (2012). Is Passenger Vehicle Incompatibility Still a Problem? *Traffic Injury Prevention*, pp. 585-591.

European New Car Assessment Programe (Euro NCAP). (2017). EuroNCAP 2025 Roadmap. Retrieved from https://www.euroncap.com/en/for-engineers/technical-papers/

Federal Highway Research Institute (BASt) and Forschungsvereinigung Automobiltechnik e.V. (FAT). (2025). *German In-Depth Accident Study Cooperation Codebook.* Retrieved 2025, from https://www.gidas.org/start-en.html

Gabler, H. C., & Hollowell, W. T. (1998). NHTSA's vehicle aggressivity and compatibility research program. *Proceedings of the 16th International Technical Conference on the Enhanced Safety of Vehicles*, pp. 640-649.

Golino, H. F., & Assis Gomes, C. (2014). Visualizing Random Forest’s Prediction Results. *Scientific Research*.

Insurance Institute for Highway Safety (IIHS). (2021). New crash tests show modest speed increases can have deadly consequences. Retrieved from https://www.iihs.org/news/detail/new-crash-tests-show-modest-speed-increases-can-have-deadly-consequences

Meghna Chakraborty, T. J. (2023). Causal Analysis and Classification of Traffic Crash Injury Severity Using Machine Learning Algorithms. *Data Science for Transportation*.

Monfort, S. S. (2025). Crash incompatibility between cars, SUVs, and pickups in 2017–2022. *Insurance Institute for Highway Safety*.

NHTSA Announces Model Year 2025 Vehicles for 5-Star Safety Ratings Testing. (2024). National Highway Traffic Safety Administration.

O'Brien, S. (2010). *Measurement and Assessment of Passenger Vehicle Compatibility in Front and Side Collisions.* Melbourne: RMIT University.

Otte, D. (2003). *Scientific Approach and Methodology of a New In-Depth-Investigation Study in Germany so called GIDAS.*

Random Forest Algorithm in Machine Learning. (2025). *GeeksforGeeks*.

The Economist. (2024). American Cars, Dangerously Heavy.

Thomas, G. E. (2006). *Compatibility and Structural Interaction in Passenger Vehicle Collisions.* RMIT University.

United States Department of Transportation. (2025). *2025 Progress Report on the National Roadway Safety Strategy.* Washington, DC: Office of the Secretary of Transportation.

Wach, W., Giddlewski , M., & Prochowski, L. (2016). Modeling Reliability of Vehicle Collision Reconstruction Based on the Law of Conservation of Momentum. *20th*

International Scientific Conference. Transport Means. Kaunas: Kaunas University of Technology.

World Health Organization. (2024). *Global Status Report on Road Safety 2023.*

Zeidler, F., & Knöichelmann, F. (1998). The Influence of Frontal Crash Test Speeds on the Compatibility of Passenger Cars in Real World Accidents. *International Journal of Crashworthiness*.

Zobel, R. (1998). *Demands for Compatibility in Passenger Vehicles.* Germany: VolksWagen AG.

ANEX I

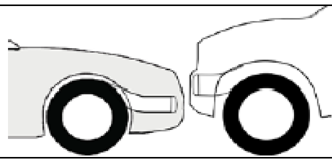
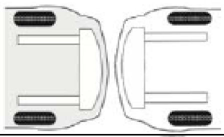
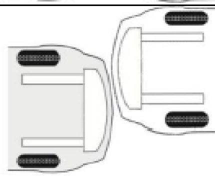
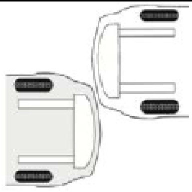
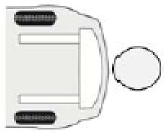
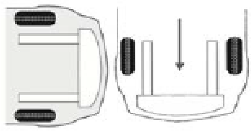
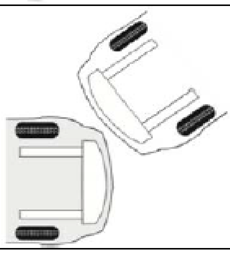
Configuration	Description	Example (case vehicle shaded)
Underride	Primary loading occurred above the bumper/longitudinal height (with any amount of horizontal overlap)	
Large overlap	Colinear loading of both longitudinals	
Moderate overlap	Colinear loading of only one longitudinal	
Small overlap	Colinear loading outboard of both longitudinals	
Center impact	Colinear loading between longitudinals	
Perpendicular	Lateral loading of both longitudinals resulting from perpendicular partner vehicle velocity	
Oblique center	Initial loading between longitudinals; often produces outboard deformation of single longitudinal	
Oblique corner	Initial loading outboard of longitudinals; often produces inboard deformation of single longitudinal	

Table 7: Crash Configurations assigned during photographic review (Brumbelow, 2019)

ANEX II

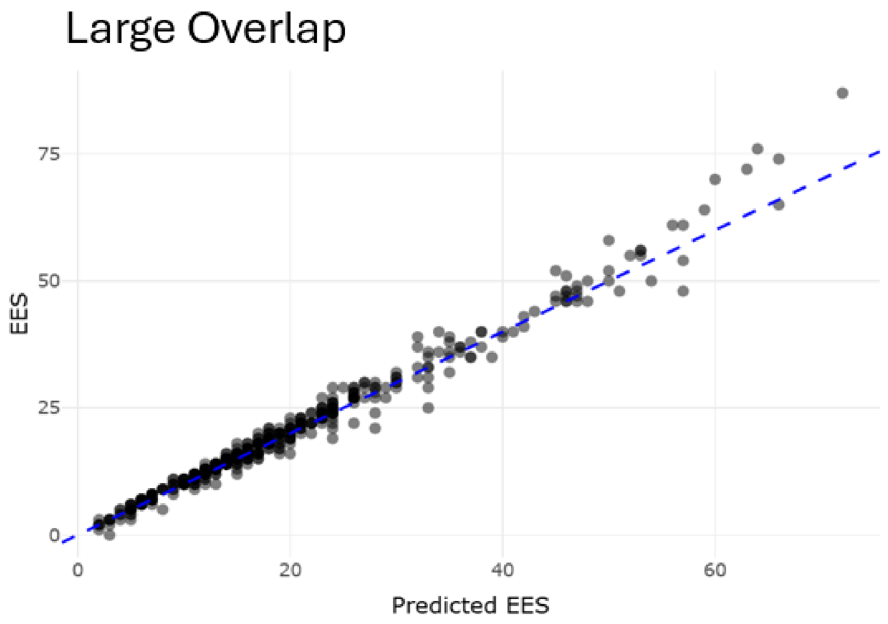


Figure 34: Comparison of Predicted EES vs. GIDAS - Large Overlap

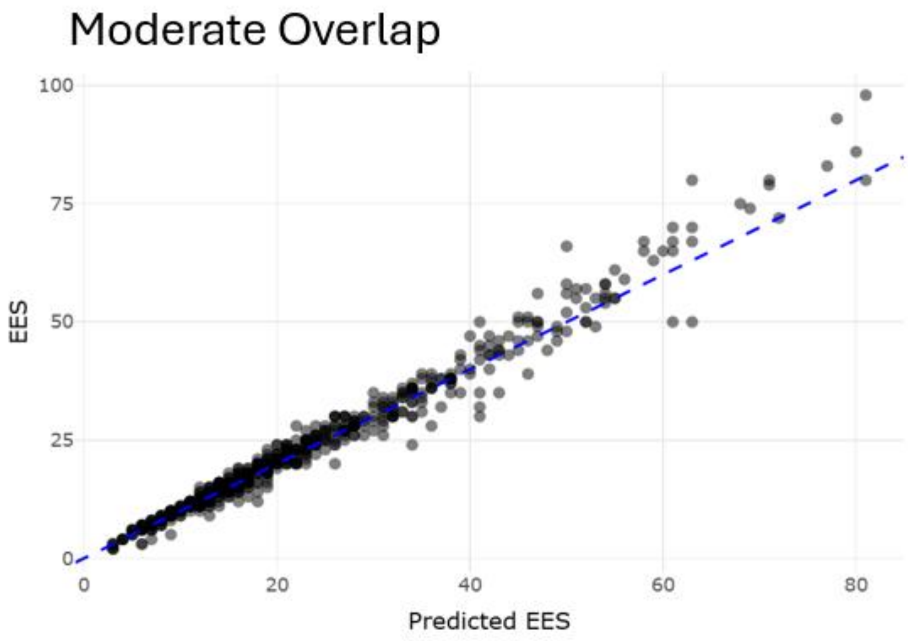


Figure 35: Comparison of Predicted EES vs. GIDAS - Moderate Overlap

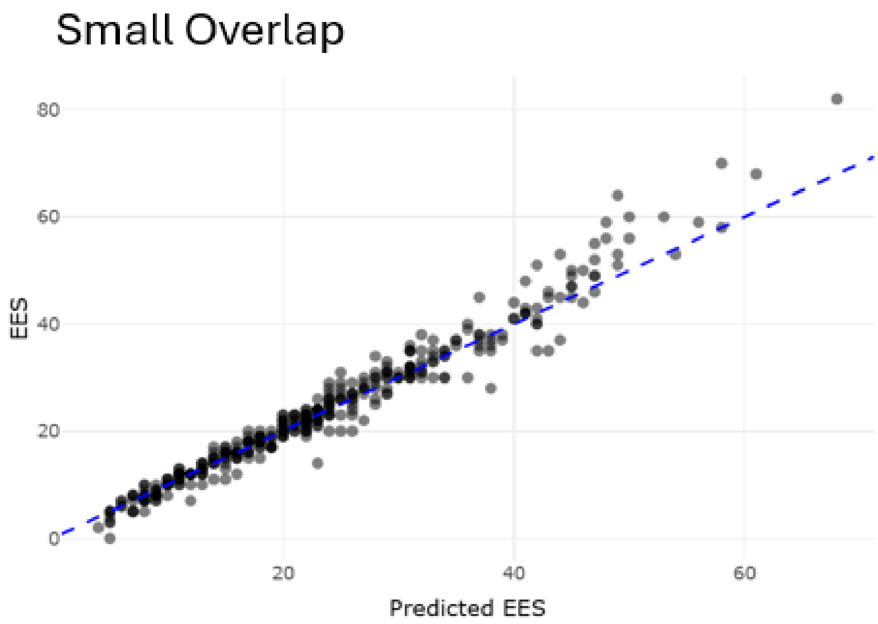


Figure 36: Comparison of Predicted EES vs. GIDAS - Small Overlap

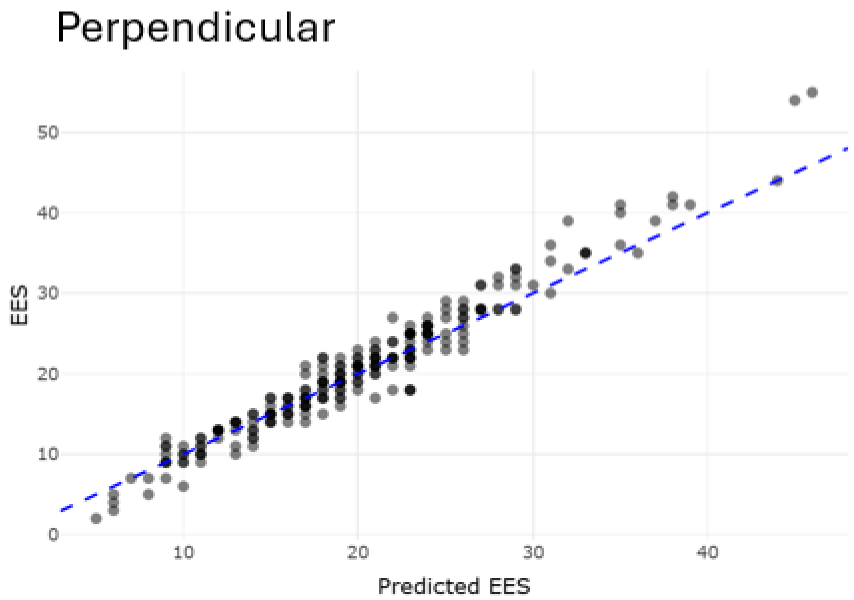


Figure 37: Comparison of Predicted EES vs. GIDAS - Perpendicular

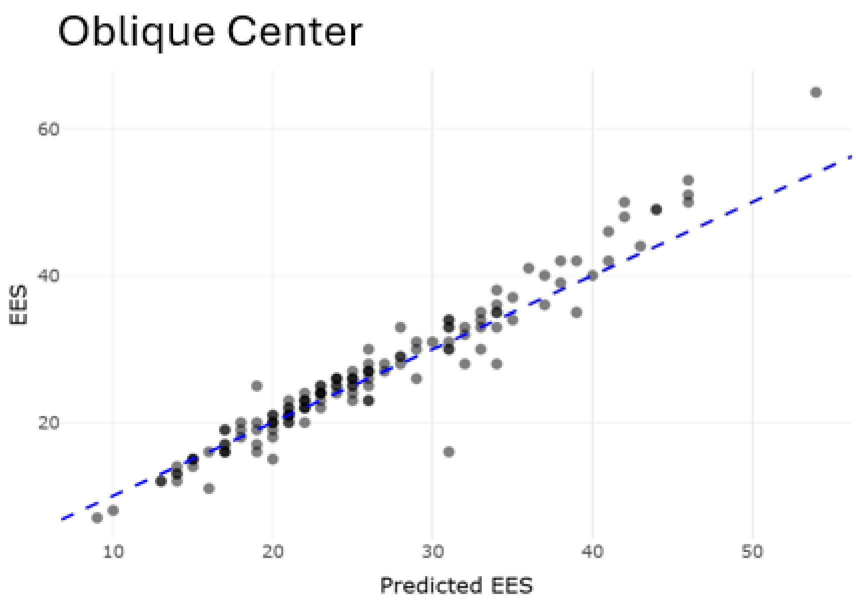


Figure 38: Comparison of Predicted EES vs. GIDAS - Oblique Center

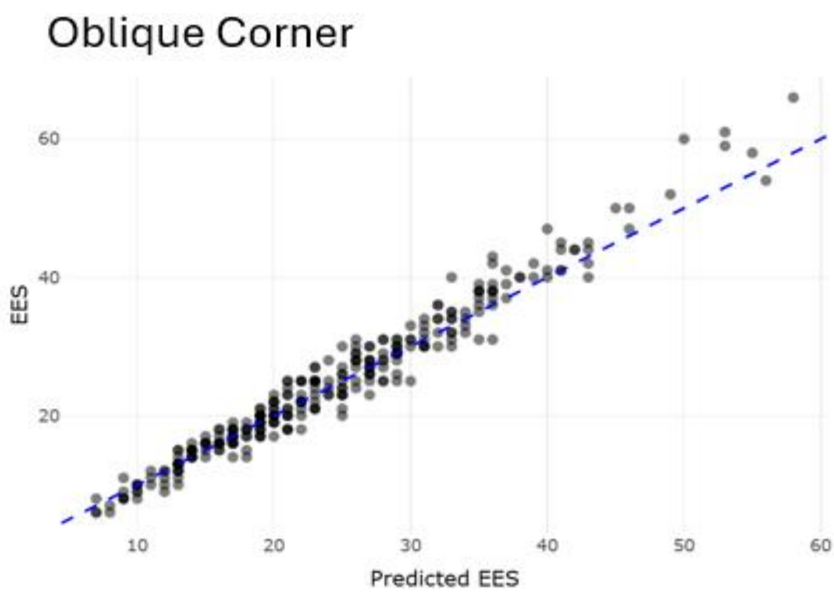


Figure 39: Comparison of Predicted EES vs. GIDAS - Oblique Corner

ANEX III

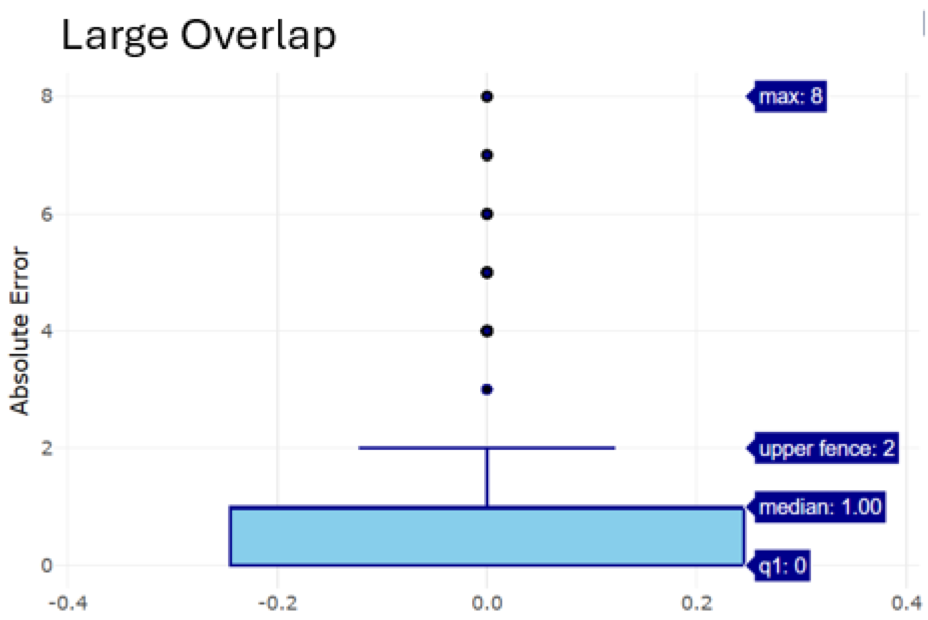


Figure 40: Error Distribution - Large Overlap

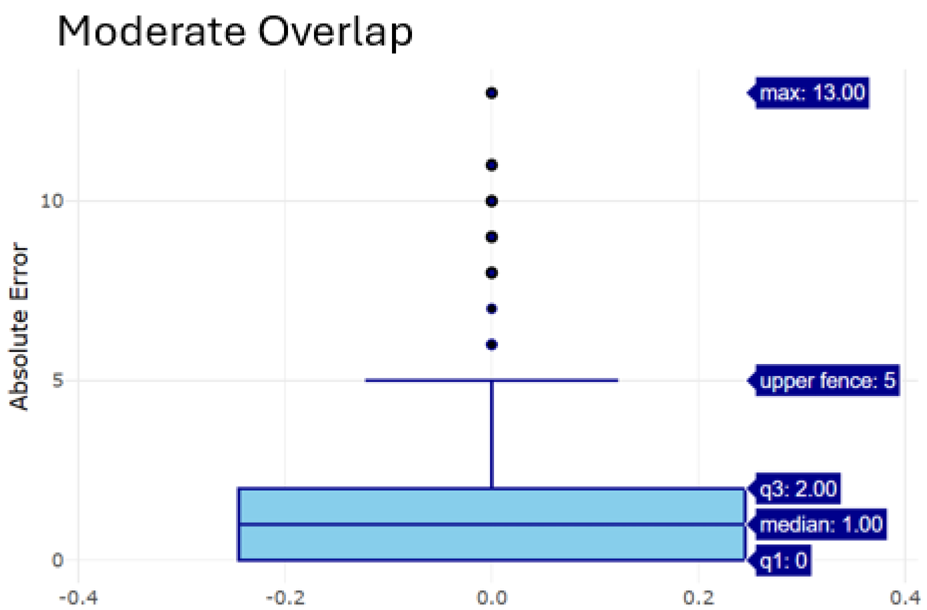


Figure 41: Error Distribution - Moderate Overlap

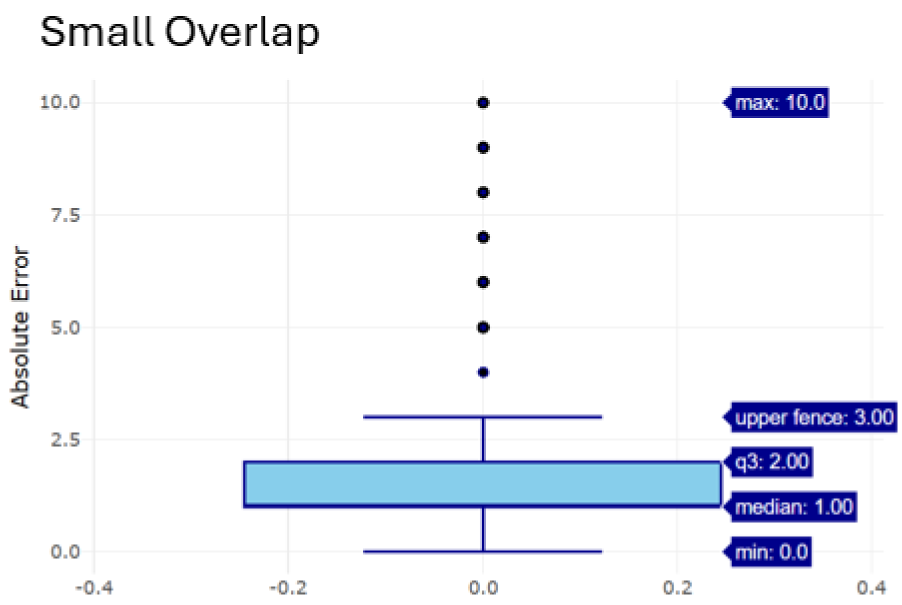


Figure 42: Error Distribution - Small Overlap

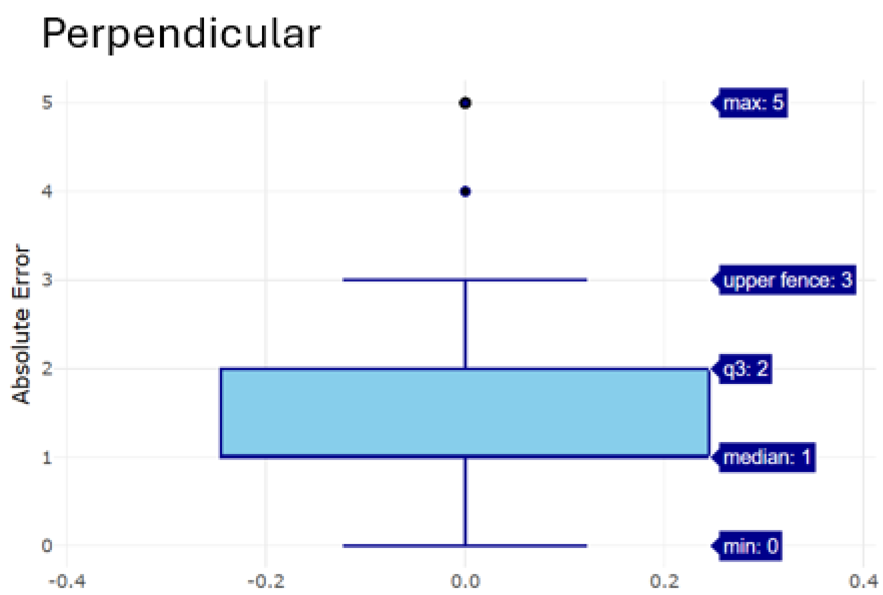


Figure 43: Error Distribution - Perpendicular

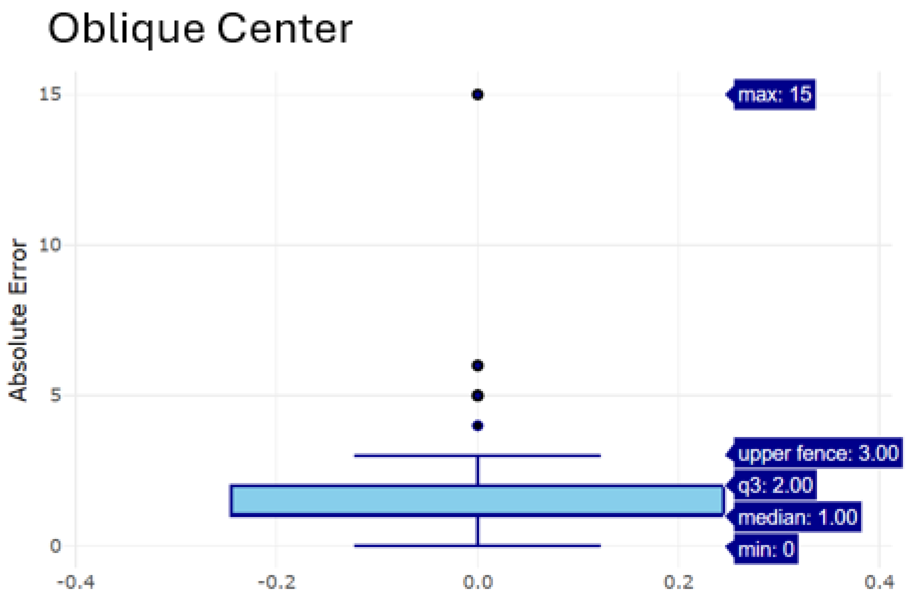


Figure 44: Error Distribution - Oblique Center

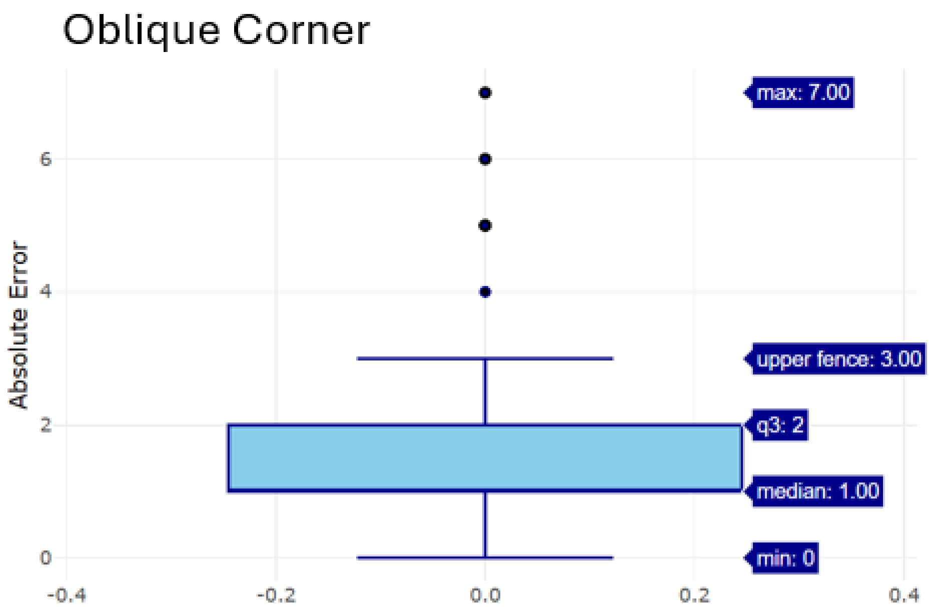


Figure 45: Error Distribution - Oblique Corner

ANEX IV

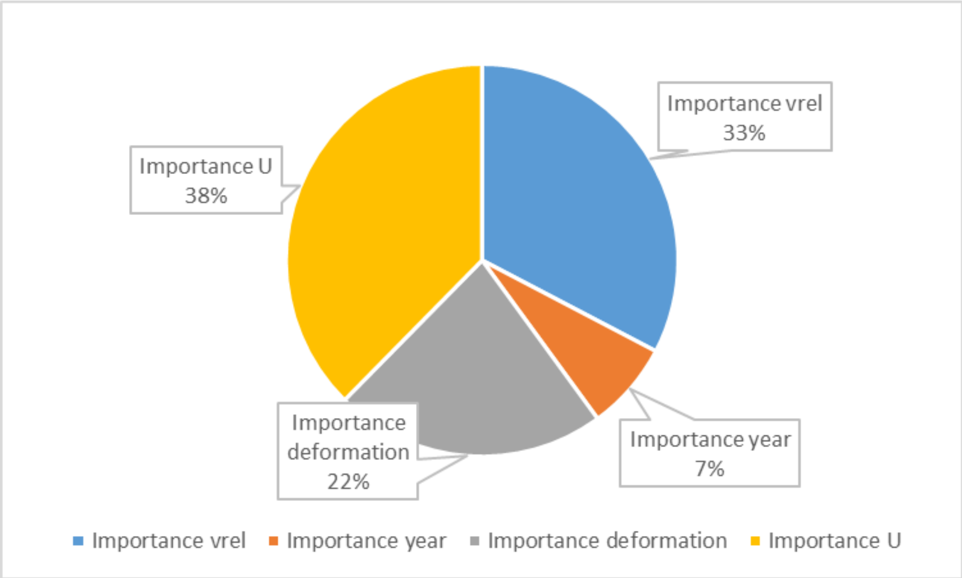


Figure 46: Importance Distribution - Large Overlap

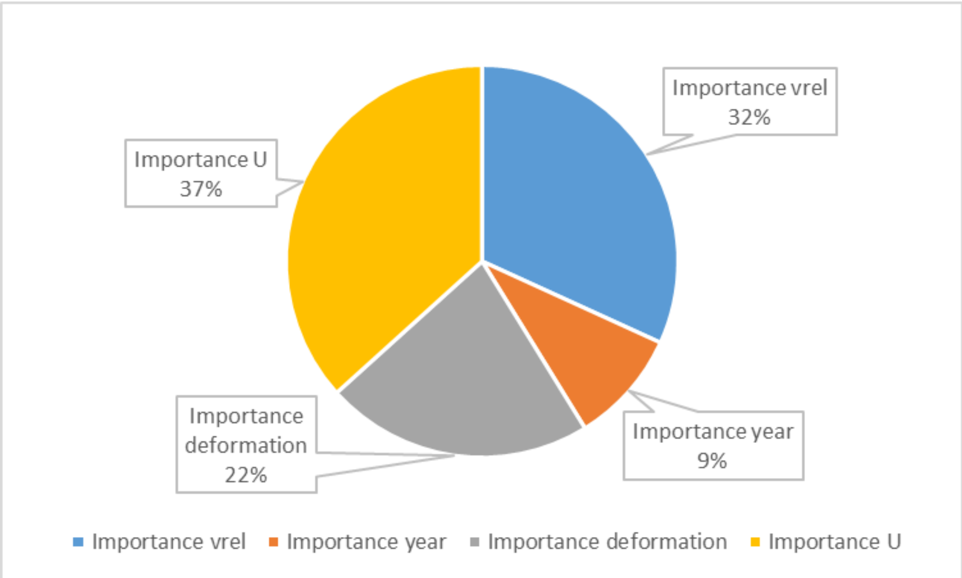


Figure 47: Importance Distribution - Moderate Overlap

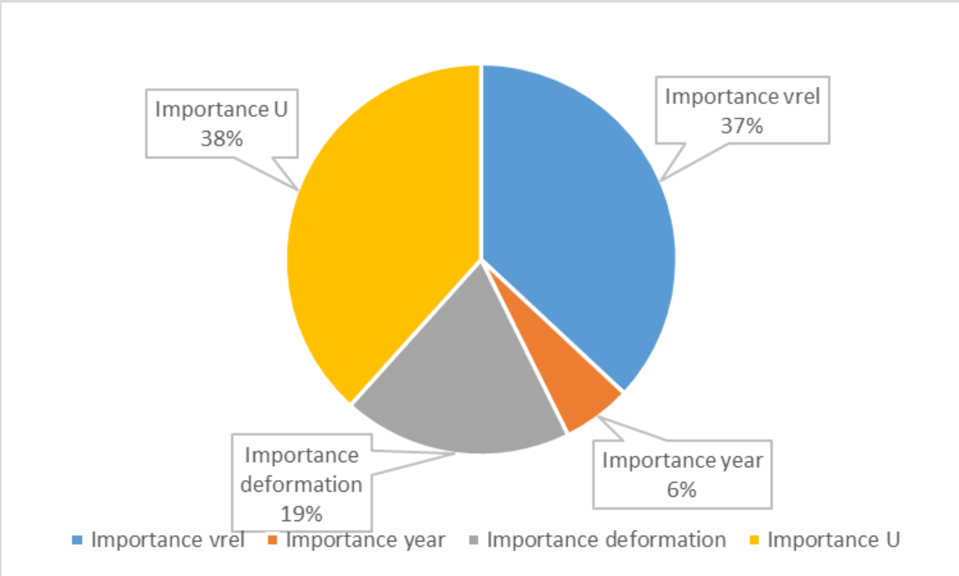


Figure 48: Importance Distribution - Small Overlap

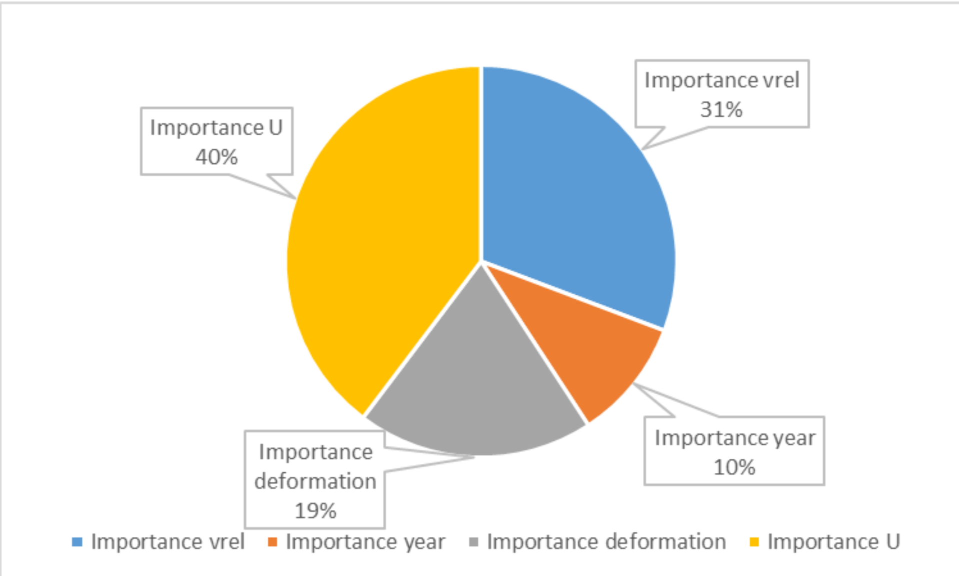


Figure 49: Importance Distribution - Perpendicular

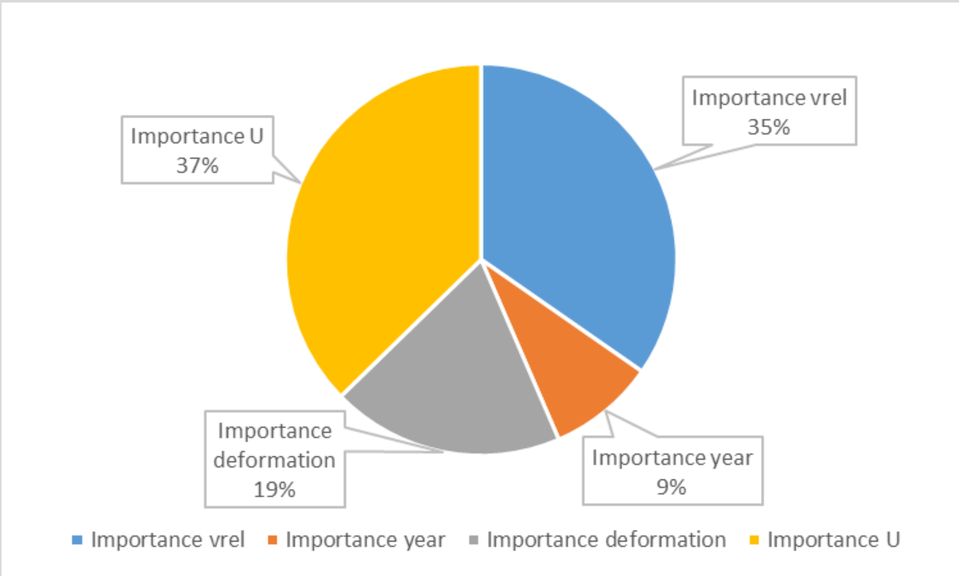


Figure 50: Importance Distribution - Oblique Center

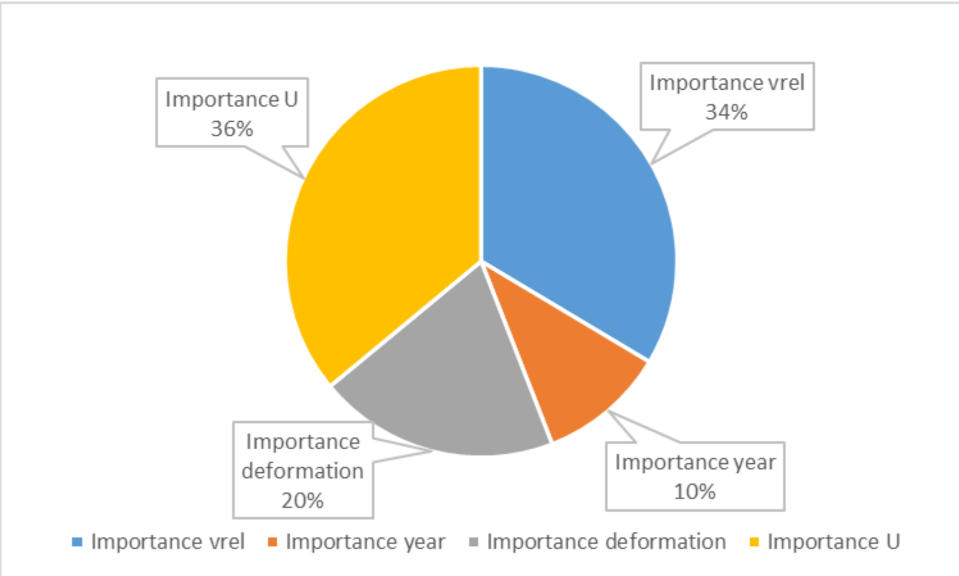


Figure 51: Importance Distribution - Oblique Corner