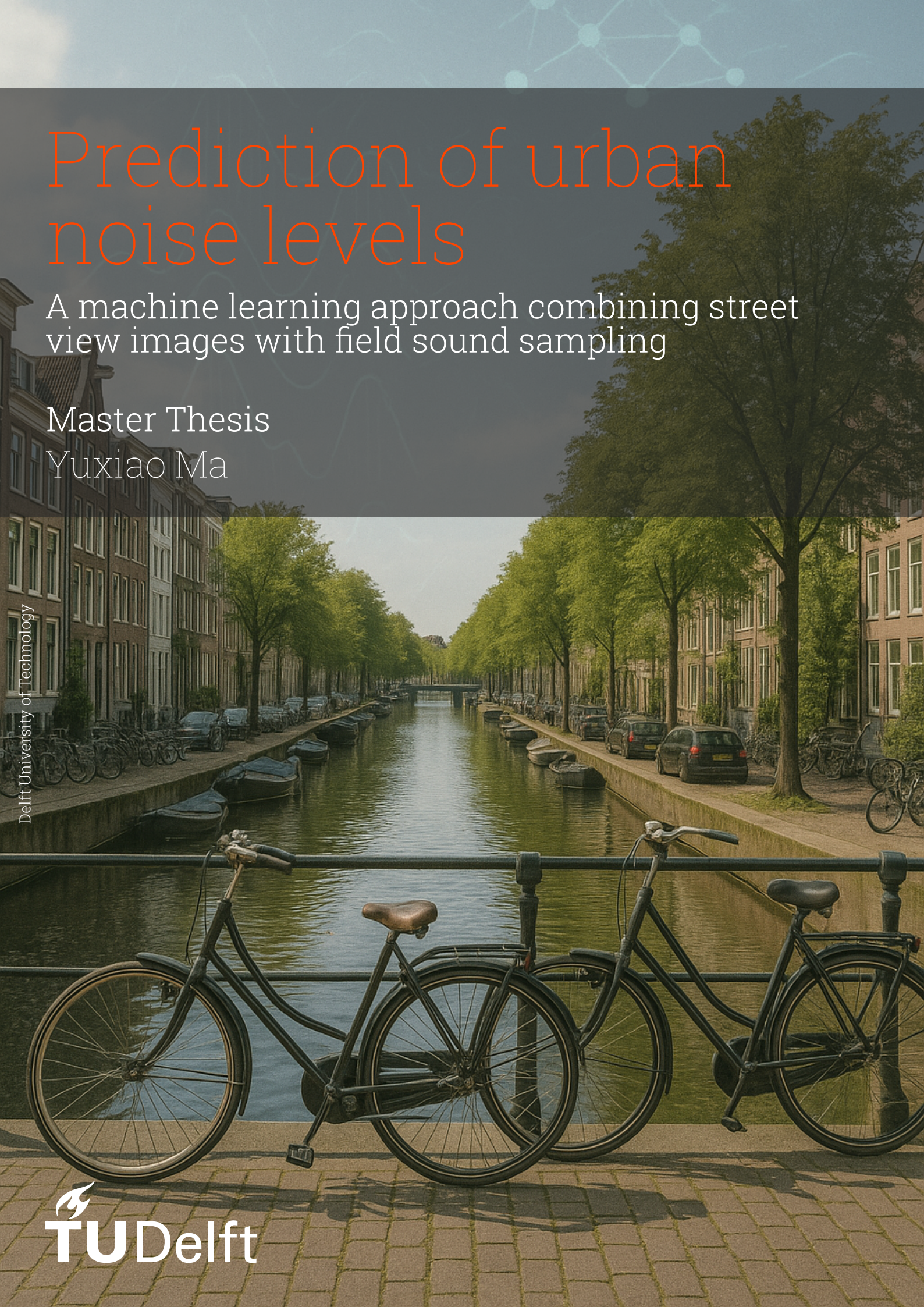
Prediction of urban noise levels

A machine learning approach combining street view images with field sound sampling

Master Thesis

Yuxiao Ma



Delft University of Technology

TU Delft

Prediction of urban noise levels

A machine learning approach combining street view images with field sound sampling

by

Yuxiao Ma

to obtain the degree of Master of Science
at the Delft University of Technology,

Student number: 5916305

Project duration: February, 2025 – July, 2025

Thesis guidance:
Sander van Cranenburgh, TU Delft, Chair supervisor
Sander Smit, TU Delft, second supervisor
Lion Cassens, TU Delft, advisor

TU Delft

Preface

This master's thesis marks the end of a challenging yet rewarding academic journey. With rapid urbanization and increasing environmental challenges, urban noise pollution has become a hidden threat to residents' health and quality of life. Traditional noise assessment methods are often expensive, have limited coverage, and struggle to capture detailed micro-environmental information. Because of this, I explored an innovative solution. I aimed to use advanced machine learning techniques, combining widely available street view image data with precise on-site acoustic measurements, to build a high-resolution, low-cost, and explainable urban noise prediction model.

During this research, I deeply learned about computer vision, acoustic analysis, and machine learning. I personally experienced the meticulous process of data collection, the complex art of feature engineering, and the repeated refinement of model tuning. From my initial curiosity about the mysteries of urban soundscapes to gradually building an intelligent system that can "recognize sound from images," every step was filled with the joy of exploration and the satisfaction of overcoming difficulties. I strongly believe that technological progress is not just about stacking numbers and code; it's about responding to real-world needs and creating better living environments for people.

The completion of this research would not have been possible without the careful guidance of my supervisors at Delft University of Technology. Their professional insights and selfless help showed me the right direction, allowing me to conduct systematic scientific exploration. I also want to thank all the teachers, classmates, and friends who provided data support, technical assistance, and encouragement during the research. Your companionship and support are the most valuable assets on my journey forward.

I sincerely hope that the findings of this thesis will provide a powerful tool for urban planners and environmental managers. This will help them understand and manage urban noise more accurately and contribute to building quieter, healthier urban environments.

Yuxiao Ma
Delft, August 2025

Abstract

Urban noise management often suffers from a gap between broad, city-level policies and the street-level conditions actually experienced by residents. This thesis develops and validates an *interpretable* machine-learning framework that combines street-view imagery (SVI) with on-site acoustic measurements to predict both a standard physical metric (A-weighted equivalent sound level, LAeq) and a psychoacoustic metric (Zwicker Loudness), which reflects how loud sounds are perceived by people. The approach integrates advanced computer-vision feature extraction with ensemble learning, and uses interpretable AI tools (e.g., SHAP) to show how specific visual characteristics of the streetscape—such as vegetation cover, road proportion, building facades, and scene perception scores—are linked to predicted noise outcomes.

Tested across several Dutch cities, the models produce consistent, street-level predictions, enabling high-resolution *diagnostic noise maps* for both LAeq and Loudness. Building on these maps, the thesis introduces a policy translation framework aligned with the Dutch *Omgevingswet* and the EU Environmental Noise Directive (END). This framework includes: (i) identifying noise "hotspots" using both absolute thresholds (from WHO guidelines) and relative, within-city rankings; (ii) diagnosing the main visual features driving noise levels, using local model explanations; (iii) selecting targeted interventions—such as traffic flow adjustments, façade and surface treatments, and nature-based solutions—supported by documented mechanisms and measurable indicators; and (iv) establishing an update loop for periodic review as new imagery and measurements become available. An illustrative micro-case demonstrates how this process turns model outputs into actionable planning decisions and performance metrics.

The study's contributions are threefold: (1) an end-to-end, interpretable pipeline linking SVI and acoustics; (2) a dual-metric evaluation (LAeq and Zwicker Loudness) that combines legal compliance with a perception-based perspective; and (3) a concrete, regulator-aligned pathway from predictions and explanations to policy action. Limitations include the sample size relative to feature dimensionality, the absence of direct GIS or morphological data integration, and the geographic focus on Dutch cities. These factors point to future work involving larger and longitudinal datasets, multi-sensor/GIS integration, and transfer learning for broader applicability, while encouraging cautious, phased adoption in real-world planning.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
ANNs	Artificial Neural Networks
CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CNN	Convolutional Neural Network
CV	Computer Vision
CV-RMSE	Cross-Validation Root Mean Square Error
DALYs	Disability-Adjusted Life Years
DCNN	Deep Convolutional Neural Network
DPIA	Data Protection Impact Assessment
END	Environmental Noise Directive
EPA	Environmental Planning Act
EU	European Union
GAM	Generalized Additive Model
GDPR	General Data Protection Regulation
GGD	Gemeentelijke Gezondheidsdienst (Municipal Health Service)
GPPs	Geluidproductieplafonds (Noise Production Ceilings)
GBRT	Gradient Boosting Regression Tree
GIS	Geographic Information System
GSV	Google Street View
HSV	Hue, Saturation, Value
IEC	International Electrotechnical Commission
IQR	Interquartile Range
ISO	International Organization for Standardization
LAeq	A-weighted equivalent continuous sound pressure level
Lden	Day-evening-night average sound level
Lnight	Night average sound level
MAE	Mean Absolute Error
MFCC	Mel-frequency Cepstral Coefficients
ML	Machine Learning
MLP	Multi-Layer Perceptron
NAPs	Noise Action Plans
NN	Neural Network
PCA	Principal Component Analysis
PCs	Principal Components
PPPs	Public-Private Partnerships
PSD	Power Spectral Density
RBF	Radial Basis Function
RF	Random Forest
RIVM	Rijksinstituut voor Volksgezondheid en Milieu (National Institute for Public Health and the Environment of the Netherlands)
RMSE	Root Mean Squared Error
RMS	Root Mean Square
RQ	Research Question
SHAP	SHapley Additive exPlanations
SPL	Sound Pressure Level

Abbreviation	Definition
SVI	Street View Images
SVR	Support Vector Regression
SWUNG	Samen Werken Uitvoering Nieuwe Geluidregelgeving (Cooperation on Implementation of New Noise Regulations)
TPE	Tree-structured Parzen Estimator
VNG	Association of Netherlands Municipalities
WHO	World Health Organization
Wm	Wet milieubeheer (Environmental Management Act)

Symbols

Symbol	Definition	Unit
N	Loudness	sone
$N'(z)$	Loudness density at a specific Bark band	
z	Bark scales	
$s_{24bit}(t)$	Scaled audio signal (in 24-bit integer range)	
$s_{norm}(t)$	Normalized input signal	
RMS	Root Mean Square	
P_{ref}	Calibration reference	
C_{cal}	Device-specific calibration parameters	
SPL_{device}	Sound pressure level estimation	dB
C_{ref}	Device calibration reference	
S_{dbfs}	Microphone sensitivity	dB
Δ_{offset}	Additional calibration offset	dB
$p_{physical}(t)$	Physical sound pressure	
x_{ij}	j-th feature value of the i-th sample	
$\bar{x}_j$	Mean of the j-th feature	
Σ	Covariance matrix	
v	Eigenvector	
λ	Eigenvalue	
Y	Projected data	
W	Projection matrix	
d	Cohen's d effect size	
s_p	Pooled standard deviation	
n_1, n_2	Sizes of the two samples	
s_1, s_2	Standard deviations of the two samples	
$\hat{y}$	Model's predicted output	
ϕ_0	Baseline value	
ϕ_j	SHAP value for the j-th feature	
M	Total number of features	
y_i	True value	
$\bar{y}$	Mean of the true values	
θ	Rotation angle	degrees
t_x, t_y	Translation values	pixels
V'	Adjusted brightness	
c	HSL channel	
μ_c	Mean of HSL channel	
σ_c	Standard deviation of HSL channel	
N	Total pixels	
p_i	Probability of scene category	
z_i	Logit score for scene category	Percentage
k	Percentage of pixels classified into class k	%

Symbol	Definition	Unit
M_i	Feature selection mask	
n_d/n_a	Feature transformer/attention dimensions	
n_{steps}	Number of decision steps	
λ_{sparse}	Sparse regularization coefficient	
lr	Learning rate	
g_i	First derivative of the loss	
h_i	Second derivative of the loss	
$\Omega(f_t)$	Regularization term	
L	Loss function	
L(t)	Objective at each step	
$p(y_i\|\theta(x_i))$	Conditional probability distribution	
x'	Scaled feature value	

1

Introduction

1.1. Urban Noise Pollution: A Growing Environmental Challenge

The 21st century has seen a sharp rise in urban noise pollution. This is a direct result of ongoing global urbanization. Mukim and Roberts (2023) predicts that about 70% of the world's population will live in cities by 2050. This shift in population will naturally increase noise pollution. It will also create a serious threat to public health and the quality of city life. Noise is an environmental stress factor. It is everywhere. Its effects are widespread and profound.

The World Health Organization (WHO) has measured this problem. They estimate that traffic noise alone causes over a million years of healthy life lost each year in Western Europe (Organization et al., 2011). At least 20% of the EU population live in areas where traffic noise levels are harmful to health (European Environment Agency, 2020). These alarming numbers highlight the hidden crisis of noise pollution. Research by Basner et al. (2014) and Hahad et al. (2019) further shows the health impacts. They linked long-term exposure to noise with many bad effects. These include mental problems like anxiety, depression, and poor thinking. They also include physical risks like more heart and blood vessel diseases. Specifically, these risks cover high blood pressure, stroke, heart attacks, and widespread sleep problems. Children are especially vulnerable to noise exposure. It can affect their thinking development. The WHO considers environmental noise to be the second greatest environmental contributor to the burden of disease in the EU after air pollution (Organization et al., 2011). In crowded countries like the Netherlands, traffic noise is the top environmental concern (National Institute for Public Health and the Environment (RIVM), 2023). It needs effective monitoring and management.

Historically, measuring noise and its complex effects has been technically hard and expensive. This study suggests new methods. These include machine learning (ML) and street view images (SVI). These can make noise assessment easier to get, more common, and more detailed. This can help people focus on noise again.

Faster urbanization and the simultaneous rise in noise pollution form a complex interaction. We can see it as a feedback loop. As city areas grow, population density increases, and economic activity strengthens. This naturally leads to more concentrated noise sources. These include noise from vehicles, construction, and daily human activities. This increased noise then harms the city's living environment and public health (Hemmat et al., 2023). The urban form itself, characterized by street canyons and hard surfaces, further intensifies the noise level through reflection and reverberation (Penteado et al., 2018). As a result, a worsening sound environment brings significant social costs. These include higher healthcare spending, lower worker productivity, and a reduced overall quality of life. This can affect the appeal and long-term sustainability of these city centers in a cycle. It also causes complex population and economic adjustments.

1.2. Problem Statement: Limitations of the Current Method and Research Gaps

Traditional urban noise evaluation methods mainly rely on manual sampling from fixed monitoring stations, mobile measurement vehicles, and handheld sound level meters (Can et al., 2011; Kirillov and Bulkin, 2015). Although these methods can provide precise measurements at specific points, they face significant challenges when trying to achieve city-wide coverage. The main issues are the high cost of equipment and maintenance, along with the need for trained personnel. As a result, large ”blind spots” are common on urban acoustic maps. A report from the European Court of Auditors (2025) also highlighted this flaw, noting that most EU member states have data gaps and delays in their noise assessments and reports, which fail to fully reflect the true noise exposure of residents.

However, an even more fundamental challenge than coverage is the widespread **”micro-macro disconnect”** in existing noise management systems. Traditional noise prediction models, such as the official EU CNOSSOS-EU model, typically operate at a macro scale, providing city- or regional-level assessments (Kephalopoulos et al., 2014). But people’s perception of noise and its health effects occur at a much finer, micro scale—on specific streets, near buildings, or even on different sides of a building.

This disconnect between the scale of assessment and the scale of impact makes it difficult for noise management to be precise and effective. Macro-level models often overlook subtle, street-level features such as building layouts, green coverage, street infrastructure, and local activity patterns, all of which can significantly alter the local acoustic environment. To bridge this ”micro-macro” gap, this study proposes an innovative approach: using street-view images (SVI) as the core data source. SVI is inherently micro-scale data that naturally captures the rich visual cues influencing the local acoustic environment. By using machine learning to analyze these images, a model can be developed to infer micro-scale acoustic environments from micro-scale visual information, thereby providing fine-grained decision-making support for macro-level policies.

1.3. Research Purpose, Objectives, and Research Questions

Purpose: This study’s main purpose is to develop and test an innovative machine learning framework. This framework serves as an **interpretable diagnostic tool** for urban noise by integrating street-view images with on-site acoustic samples. The purpose is to provide **high-resolution, actionable insights** that can bridge the existing ”micro-macro” gap in noise management and support more precise urban planning.

Research Objectives To achieve the main purpose, this research focuses on the following specific objectives:

1. **Quantify the visual-acoustic correlation**: This involves investigating and quantifying the relationship between visual features in street-level images and objective environmental noise levels. The goal is to identify the most significant visual predictors of urban noise.
2. **Develop an interpretable predictive model**: A machine learning model is constructed and evaluated. This model is designed not only to predict noise levels with practical accuracy (including A-weighted equivalent continuous sound pressure level, LAeq, and psychoacoustic metrics of human perception) but also to provide a transparent explanation of why specific visual features lead to corresponding acoustic outcomes.
3. **Establish a policy translation framework**: A practical framework is developed to translate the model’s outputs (high-resolution noise maps and feature explanations) into concrete policy interventions. This framework is designed to be consistent with the regulatory contexts of both the Netherlands (Omgevingswet) and the European Union (END).

Research Questions Based on the above background and objectives, this research raises the following core questions:

> *How can an interpretable SVI–acoustics framework be built to predict urban noise and provide actionable support for planning and management?*

In order to fully answer this central question, this study will also explore the following sub-questions:

- **RQ1:** *Which quantifiable visual characteristics of an urban streetscape does a machine learning model identify as the most significant predictors for both physical noise levels (LAeq) and perceived psychoacoustic loudness (Zwicker Loudness)?*
- **RQ2:** *How can the model be structured and explained so that its predictions are not only accurate but also accompanied by clear, credible reasons for why specific visual features drive the outcome?*
- **RQ3:** *How to transform model outputs into concrete and actionable noise control intervention plans that are aligned with the Dutch Omgevingswet and EU END indicator systems?*

1.4. Research Scope and Limitations

Geographic Scope: The primary data for model training and initial testing were collected in The Hague and Delft, Netherlands. The model's generalizability will be explored by applying public Google Street View images from other Dutch cities, such as Amsterdam, Rotterdam, etc.

Methodological Scope: This study focuses on applying machine learning techniques, specifically those capable of integrating visual features from street-view images (SVI) with on-site, short-term acoustic measurements.

Limitations: To maintain a focused and feasible study within the timeframe of a master's thesis, the following boundaries were explicitly set:

Data Sparsity Challenge

This study faced the typical challenge of a high-dimensional feature space (over 350 raw visual features) with a limited sample size (approximately 1400 valid data points). This is not just a limitation but a core finding of the study. A key objective was to explore the predictive limits of this method under real-world data constraints and to establish a performance benchmark for future large-scale research.

No Integration of GIS Data

This was a deliberate boundary. The study did not directly integrate traditional GIS datasets (such as 3D building models, detailed road network attributes, or land use layers). This choice was made to purely test the capacity of street-view images to represent environmental information and to answer the fundamental question: "To what extent can we understand the acoustic environment by merely 'looking' at the street?" While this sharpened the research focus, it also means the model lacks the macro-spatial information that GIS data could provide and which might improve physical accuracy.

Limited Psychoacoustic Depth

Although key psychoacoustic metrics (such as Zwicker loudness) were introduced, the study did not cover more comprehensive parameters (such as sharpness or roughness). Consequently, the simulation of human noise perception is considered preliminary.

Geographic Generalizability

The model was primarily trained and validated within a Dutch urban environment, making it, in essence, a "Dutch urban noise expert." Its direct applicability to other international cities with vastly different urban morphologies, architectural styles, or traffic cultures would require further investigation through transfer learning or recalibration with local data.

1.5. Thesis Structure

This thesis moves from **understanding data** to **building models** and then to **applications and discussion**. Chapter 5 closes by extracting actionable rules and explicitly sets up the policy translation in Chapter 6. Chapter 7 revisits the initial motivation (the micro–macro disconnect) in light of the results, consolidating contributions and limitations. This structure clearly shows the whole research process.

Chapter 1: Introduction
This chapter discusses the severity of urban noise pollution, highlights the limitations of existing noise assessment methods, and introduces the core concept of this study. Finally, it clearly states the research objectives, new research questions, and the scope and boundaries of the research.

Chapter 2: Theoretical Background and Literature Review
This chapter provides a solid theoretical foundation for the study. It reviews the impact of urban noise on health and society, traditional and emerging noise assessment methods, the role of psychoacoustics in understanding noise perception, and relevant noise management policies in the Netherlands and the European Union.

Chapter 3: Research Methodology
This chapter details the technical workflow of the study. It includes the methods for collecting acoustic and visual data, feature engineering techniques used to quantify the visual environment, the selection of machine learning models and their interpretability frameworks, and the policy translation framework designed to address Research Question 3 (RQ3).

Chapter 4: Acoustic Data Exploration and Feature Analysis
Before building the predictive model, this chapter provides an in-depth exploration of the collected data. The core content validates whether different urban soundscapes have unique and identifiable ”acoustic fingerprints,” laying the data foundation for the subsequent cross-modal prediction task.

Chapter 5: Predictive Modeling: Results and Interpretation
This chapter presents the core results of the predictive model. The content directly addresses RQ1 by identifying key visual predictors through global feature importance analysis. It also addresses RQ2 by demonstrating the model's interpretability through SHAP analysis on specific samples.

Chapter 6: Discussion and Policy Implications
This chapter provides an in-depth discussion of the modeling results, examining their scientific and practical value. The content systematically addresses RQ3, exploring how to translate the model's technical outputs into a practical policy recommendation framework for urban planning and environmental management.

Chapter 7: Conclusion and Future Research
This chapter summarizes the thesis, reiterates the core contributions of the research, and proposes future research directions based on the challenges discovered in this study.

2

Theoretical Background and Literature Review

This chapter lays the groundwork for the research. It reviews existing knowledge in several key areas. These include: the wide-ranging effects of urban noise, traditional and modern noise assessment methods, the role of psychoacoustics in understanding noise perception, and the relevant policies and rules for managing noise in the Netherlands and the European Union.

2.1. Urban Noise: Health and Social Impact

2.1.1. Physical and mental health effects

Urban noise is more than just an annoyance. It is a major environmental stressor. It has profound and well-documented impacts on human health and well-being. Long-term exposure to high noise levels is very common in many city environments. This exposure can trigger a range of negative physical and psychological reactions.

Cardiovascular health is significantly affected. Münzel et al. (2018) conducted a study and they concluded that exposure to environmental noise (and other environmental stressors) can lead to cardiovascular metabolic diseases. Noise can cause oxidative stress, problems with blood vessel function, imbalances in the autonomic nervous system, and metabolic issues. This can worsen the negative health effects of traditional risk factors like high blood pressure, diabetes, and high cholesterol. For example, it can speed up the progress of atherosclerosis and increase the risk of cardiovascular events.

For sleep and mental health, Smith et al. (2022) showed there's moderate evidence for the probability that for every 10 dB increase in nighttime noise, quality of sleep will be highly disturbed. This then leads to chronic stress and anxiety. National Institute for Public Health and the Environment (RIVM) (2023) claims that in the Netherlands, road traffic and neighbors are the main sources of noise disturbance. They are also the primary causes of sleep problems. More than 4% of the adult population in the Netherlands experiences severe sleep disturbances due to road traffic noise. Like noise disturbance, the percentage of severe sleep disturbances can vary greatly locally compared to the national average. For example, around Schiphol Airport, an average of 4% of residents suffer from severe sleep disturbances due to air traffic (Oosterlee and Zandt, 2017).

From a psychological perspective, chronic noise exposure is linked to more stress, anxiety, depression, and mood disorders. Cognitive functions, such as attention, memory, and the ability to learn, can also suffer. Children are especially vulnerable in this area. Klatte et al. (2013) reviewed studies showing that noise in learning environments hinders speech perception, reading comprehension, and memory development in school-aged children. The WHO (Organization et al., 2011) also measured the burden of disease caused by environmental noise. For child cognitive impairment, they estimated 45,000 disability-adjusted life years (DALYs) are lost each year in high-income Western European countries for this specific outcome. Older adults are another sensitive group. Studies show an increased stroke

risk linked to higher noise exposure. These widespread effects impact the physical and psychological health of different age groups. This requires a strong understanding of noise and effective ways to reduce it. Therefore, our proposed model can identify noise hotspots. This directly connects to public health goals aimed at protecting these vulnerable groups. This strengthens the reasoning for RQ3, which focuses on policy changes.

2.1.2. Socioeconomic and Environmental Justice Impacts

Noise pollution has significant economic consequences. The noise pollution cost in the European Union includes healthcare expenses, lost productivity, and reduced property values. A joint report from the French Agency for Ecological Transition (ADEME) and the French National Noise Council (CNB), estimates that noise pollution costs France **147.1 billion** euros annually (ADEME and Conseil National du Bruit, 2021). Transport noise makes up the largest part of this, accounting for 66.5% (97.8 billion euros) of the total. Organizations like I Care & Consult and Energies Demain conducted this study. They looked at how different noise sources, including transport, communities, and workplaces, affect health and the economy. The report also notes that 86% of these costs are non-market costs, like health problems and a lower quality of life. The remaining 14% are direct market costs, such as medical expenses, lost productivity, and lower property values.

A report from the Dutch National Institute for Public Health and the Environment (RIVM) states that property depreciation due to airport noise in the Netherlands is estimated at about 1 billion euros annually (Schreurs et al., 2011). There is also a 400 million euro reduction in land value, making the total loss around 1.4 billion euros. Amsterdam Airport Schiphol accounts for about 65% of these costs. Another study by RIVM, in cooperation with EFTEC, estimates the median annual social cost of noise pollution in the Netherlands to be 1.81 billion euros (Howarth et al., 2001). This mainly includes lost productivity, lower property values, and a decreased quality of life.[1]

For environmental justice, noise pollution makes social inequality worse. Casey et al. (2017) found that noise pollution is not spread equally among people from different racial/ethnic and socioeconomic backgrounds. Several indicators of a community's socioeconomic background are linked to more noise at night and during the day. These include poverty, unemployment, language isolation, and a higher percentage of renters and people who did not finish high school. Neighborhoods with more Asian, Black, and Hispanic residents often have higher noise levels, but these relationships are rarely linear. Dutch noise maps (Atlas Leefomgeving, 2025) also show that immigrant communities in The Hague, like Transvaal, experience more traffic noise than wealthier areas like Wassenaar.

2.2. Conventional Noise Assessment: Advances and Challenges

Traditionally, urban noise assessment combines direct measurement and simulation modeling. Direct measurement involves placing fixed monitoring stations in key spots. It also uses mobile surveys with vehicles that have advanced acoustic tools. Manual on-site measurements are also done with handheld sound level meters. These methods provide very accurate Leq values and other acoustic data at the measurement points. Computational simulation models, like the European common noise assessment method (CNOSSOS-EU) framework, predict noise levels. They use inputs such as traffic flow data, road characteristics, building shapes (often from GIS), and weather conditions (Kephalopoulos et al., 2014).

These traditional methods are well-established, but they have significant limitations. This is especially true when a city needs a comprehensive, high-resolution assessment. Key limitations include:

Cost and Scalability: Setting up and maintaining a dense network of high-quality fixed monitoring stations is very expensive. The capital and operational costs for frequent, full mobile surveys are also substantial. Expanding this to new cities or regions means a large, long-term budget commitment for setup and upkeep.

Spatial Coverage: Due to cost limitations, monitoring points are often spread out. This leaves large unmonitored areas and an incomplete picture of the urban soundscape. It makes it hard to identify

[1] It is important to note that this estimate is primarily based on data from 2000, so it might not fully reflect the current situation.

localized noise hotspots or assess exposure in different small environments.

Micro-environment Details: Large-scale simulation models are useful for strategic mapping at a **macro-scale**, however, they often simplify complex urban structures. They may not fully capture how **micro-scale** features like specific building facades, small green spaces, or street furniture affect local sound spread.

Data Intensity of Physical Models: Advanced physical simulation models, like CNOSSOS-EU, are ”data-hungry”. They need extensive and current input data on traffic, 3D building shapes, terrain, and weather. Getting and maintaining these detailed datasets is a major challenge and expense for many cities, especially smaller ones. Any inaccuracies or outdated information in the input data directly affect the reliability of the simulation output. This need for data contrasts with the potential of ML-SVI methods. These methods can provide insights even when data is scarce. They can also offer quick updates between major regulatory modeling cycles, making them potentially more cost-effective.

These limitations highlight the need for methods that can provide more detailed, cost-effective, and dynamic noise assessments.

2.3. Progress in Noise Prediction: Machine Learning and New Data Sources

To address the shortcomings of traditional **macro-scale** methods, researchers have turned to innovative approaches that leverage machine learning (ML) and new data sources, which are particularly adept at capturing environmental characteristics at the **micro-scale**. Among these, Street View Imagery (SVI) has emerged as a powerful resource.

2.3.1. Street View Images (SVI) for Urban Environment Analysis

Street View Images (SVI) platforms, like Google Street View, offer a vast and easily accessible source of visual information about city environments. These images capture details such as road width, building density and type, vegetation cover, traffic presence, and other street-level elements. All of these are inherently linked to local noise levels. Using computer vision and ML with SVI opens new possibilities for understanding the environment.

Several studies have shown the potential of SVI in noise and soundscape assessment:

Zhao et al. (2023) successfully used SVI to estimate the urban soundscape of large areas in Singapore and Shenzhen with high resolution. Their method involved extracting visual features from SVI, conducting surveys to gather soundscape indicators based on these images, and then training a **Gradient Boosting Regression Tree (GBRT)** model. They achieved a coefficient of determination (R^2) of 0.48 when validating predicted sound intensity against on-site measurements, which shows that SVI-based soundscape sensing is feasible.

Huang et al. (2024) proposed a hybrid method. It combined a **Deep Convolutional Neural Network (DCNN)** to extract visual features from SVI with traditional machine learning algorithms (Random Forest) to predict road noise levels. Their work highlights the ability to learn meaningful patterns from visual data for quantitative noise estimation, achieving a Mean Absolute Error (MAE) of 2.01 dB and a Root Mean Square Error (RMSE) of 2.71 dB in their study.

Verma et al. (2020) combined locally acquired SVI and synchronized audio recordings to predict urban environment perception. Their visual feature extraction methods, which used tools like **Faster R-CNN** for object detection, **PSPNet** for semantic segmentation, and **Places365-trained CNNs** for scene classification, set a strong precedent for the feature engineering aspect of this paper.

Song et al. (2024) introduced a multi-sensory framework. It used models like **XGBoost (or LightGBM)** and **SHAP** for explainable noise perception classification based on SVI. A key finding was the significant impact of visual scene composition (the proportion of buildings, vegetation, sky, and lighting) on perceived noise, further validating the use of image data.

The advantages of ML-SVI methods are many. They include the potential for **high spatial resolution mapping, cost-effectiveness** compared to traditional widespread monitoring, the ability to **combine information from multiple sources** (visual and acoustic), **predictive capabilities for scene simulation**, and enhanced model interpretability through appropriate techniques.

Combining visual and acoustic data is a key part of modern noise prediction research. There are different ways to combine this data. These include early fusion, which combines features before or at the input layer of a model, and late fusion, which combines the outputs of separate models trained on single types of data.

Hong et al. (2020) gave a good example of audio-visual fusion. They trained separate **convolutional neural networks (CNNs)** on images and audio spectrograms. Then, they used a late fusion method with a linear model or a **generalized additive model (GAM)** for the final noise mapping. Their model was very accurate (R^2 > 0.90 for 5-minute average noise levels). Importantly, they found that audio features were often better at predicting environmental variables like noise than image features alone. This shows how important it is to include acoustic data in training, even if the final goal is to predict only from Street View Imagery (SVI).

However, multimodal methods have challenges. The ”modality gap” refers to the difficulty of using a model trained on rich audio-visual data for predictions where only visual input, such as Google Street View (GSV) images, is available. ”Temporal mismatch” is another concern. This happens when outdated SVI doesn't match current noise conditions.

When applying machine learning (ML) to environmental modeling for policy, a key consideration is balancing model accuracy and how easy it is to understand. Very complex models like deep CNNs, such as those Hong et al. (2020) used, might offer better prediction accuracy. However, simpler models, like rule-based systems or models using ”concept bottlenecks” (e.g., using object counts as understandable intermediate features), offer more transparency. Policymakers often need to know why a model made a certain prediction to build trust and design effective interventions. ”Black box” models can make this understanding difficult.

2.4. Psychoacoustics: Understanding Human Perception of Noise

While objective physical measurements like L_{eq} (equivalent continuous sound level) are crucial for quantifying noise, human perception of sound is a more complex phenomenon. It is affected by various psychoacoustic properties of the sound itself. Understanding these properties is vital for designing interventions that not only lower decibel levels but also improve the perceived quality of the acoustic environment and minimize annoyance (Sustainability Directory, 2025). Key psychoacoustic parameters include:

Loudness: This is the subjective perception of sound intensity. It isn't linearly related to sound pressure level (dB). Measures like Phons and Sones are used to quantify perceived loudness. **Zwicker loudness**, an international standard (ISO532-1:2017, 2017), measured in **sones**, developed by Eberhard Zwicker and his team (Zwicker and Fastl, 2013), is an international standard for calculating how loud a sound is. It reflects how the human ear subjectively perceives loudness more accurate ly than traditional sound measurement methods like L_{eq}.

This model fully considers how the human ear is more sensitive to some sound frequencies than others. It also accounts for how louder sounds can mask quieter ones, and that our perception of loudness doesn't increase linearly with sound intensity. For example, **1 sone** is defined as the loudness of a 1 kHz pure tone at **40 dB SPL**. If the loudness doubles (like from 1 sone to 2 sones), it means the sound subjectively feels twice as loud.

Because of this, even if two noises have the same L_{eq} value, their Zwicker loudness can be very different. This depends on their **spectrum** (the mix of frequencies) and how they change over time. The Zwicker loudness usually better shows how annoying a sound truly is.

Using Zwicker loudness in **urban noise research and management** allows for a more complete evaluation of sound environment quality. It helps identify noise sources that might not exceed limits but

still cause significant disturbance. This also provides a more human-centered basis for urban planning and noise control.

For instance, a steady, low-frequency rumble might be less annoying than an intermittent, high-pitched whine, even if both have the same L_{eq}. This qualitative aspect of noise annoyance is critical. The human auditory system isn't just a simple microphone; it processes sounds through complex filtering and interpretation mechanisms in the brain. Therefore, aligning noise assessment with human cognition, as requested, requires a thorough understanding of these psychoacoustic dimensions.

While current models mainly predict L_{eq}, any discussion about their credibility and alignment with human perception must include these insights. Future model iterations could aim to directly predict these more nuanced metrics or use them to weight or explain L_{eq} maps, leading to a more comprehensive understanding of the urban sound environment.

2.5. Noise Management Policies and Regulatory Landscape in the Netherlands and the European Union

This research fits within the strong policy and regulatory frameworks for noise management at both the EU and Dutch national levels. Key legal tools include:

The EU Environmental Noise Directive (END) 2002/49/EC: This directive is the foundation of EU noise policy (European Parliament and Council of the European Union, 2002). It requires member states to create strategic noise maps every five years for major urban areas (over 100,000 residents), main roads (over 3 million vehicles per year), main railways (over 30,000 train passages per year), and major airports (over 50,000 movements per year). These maps must use consistent noise indicators, mainly L_{den} (day-evening-night average sound level, with penalties for evening and night) and L_{night} (night average sound level). The END also requires Noise Action Plans (NAPs). These plans aim to manage noise problems and protect existing quiet areas. While the END does not set mandatory EU-wide limits, L_{den} > 55 dB and L_{night} > 50 dB are often used as thresholds for reporting exposed populations and triggering action.

The Dutch Environmental Management Act (Wet milieubeheer - Wm): This act is the main national tool for implementing the END in the Netherlands. It provides the legal basis for controlling noise pollution, noise zoning, and enforcing legal noise standards. For example, it sets preferred maximum noise levels for residential areas (usually around L_{den} 50-55 dB(A)). It also established a system of noise "production caps" (Geluidproductieplafonds - GPPs) for major infrastructure like roads and railways, especially under the SWUNG (Samen Werken Uitvoering Nieuwe Geluidregelgeving) amendment (Grantham Research Institute, 2004).

The Dutch Environment and Planning Act (Omgevingswet) (effective January 1, 2024): This landmark law combines and simplifies over twenty existing environmental and planning laws, including the previous Noise Nuisance Act (Wet geluidhinder). Its main goal is to promote integrated decision-making and give significant planning authority to municipalities. A key tool under the Environment Act is the municipal environmental plan (omgevingsplan). This plan replaces old land-use plans and integrates rules for the physical living environment, including noise. A core principle of the Environment Act is "evenwichtige toedeling van functies aan locaties" (balanced allocation of functions to locations). This clearly requires environmental factors like noise to be considered in spatial planning decisions. The Besluit kwaliteit leefomgeving (Bkl - Decree on the Quality of the Living Environment) sets national rules that municipal environmental plans must follow. This includes a framework for noise standards (standardwaarden - standard values, and grenswaarden - limit values) for noise-sensitive buildings and areas (Grantham Research Institute, 2024).

The Municipal Public Health Service (Gemeentelijke Gezondheidsdienst (GGD)) provides health-based recommendations for evaluating environmental noise exposure. These guidelines are used to assess the potential health risks posed by various noise sources such as road traffic, railways, industrial activities, and air traffic. The GGD emphasizes these values serve as key reference points for urban planning, environmental assessment, and public health evaluations (Rijksinstituut voor Volksgezondheid en Milieu (RIVM) and Gemeentelijke Gezondheidsdiensten (GGD)'en, 2019). The summary of these noise level thresholds is presented in Table 2.1.

Table 2.1: GGD Health-Based Noise Guidelines by Source Type

Noise Source	**Day-Evening-Night Level L_{den} (dB)**	**Night Level L_{night} (dB)**	**Notes**
Road, rail, and industrial noise	50 (recommended maximum)	40 (recommended maximum)	Protects health and sleep
Aircraft noise	45 (WHO guideline)	40 (WHO guideline)	More disturbing than other sources; higher health impact
Indoor noise (any source)	33 (maximum L_{den})	–	For a healthy indoor environment

The introduction of the Omgevingswet is especially important as it creates a strong need for detailed, localized environmental data, which the proposed ML-SVI model is well-positioned to provide.

In summary, this literature review reveals several critical gaps in current urban noise research. First, the limitations of traditional assessment methods in cost and spatial resolution have led to a persistent **disconnect between macro-scale policy and micro-scale soundscape experience**. Second, while emerging research using SVI and machine learning shows promise, many studies still focus on single physical metrics and often lack sufficient **model interpretability** to be trusted as ”diagnostic tools” by policymakers. Third, the discrepancy between physical noise (e.g., LAeq) and human perception (e.g., loudness), though well-established in psychoacoustics, is rarely addressed by models that can predict and utilize both metrics for a **multi-dimensional soundscape quality assessment**. Finally, a systematic framework for **translating** the outputs of these advanced models into **practical policy actions** that align with existing regulations, such as the Dutch Omgevingswet, remains a largely unexplored area.

To address these challenges, this thesis poses its central research questions, aiming to fill these gaps with an interpretable, multi-modal machine learning framework.

3

Research Methodology

3.1. Research Design Overview

This chapter describes the technical framework for answering the study's three main research questions. The framework was a multi-stage, goal-oriented research process that required the development of an accurate and explainable high-resolution urban noise prediction model. This model was developed by integrating street-view images (SVI) with on-site acoustic measurements to provide a scientific basis for urban planning and environmental policy.

The overall process of this study followed a path from data to insights to application. It began with the **data collection and preparation phase**, where a multi-modal dataset covering several cities in the Netherlands was built using both on-site sampling and public data platforms. Next, during the **feature engineering phase**, advanced computer vision techniques were used to convert raw, unstructured visual data (images) into structured numerical feature vectors with over 350 dimensions that a machine learning model could understand.

Before proceeding with cross-modal prediction, an important prerequisite was to confirm that the acoustic data used as the prediction target contained stable, effective, and distinct signals. This soundscape exploration and validation analysis is detailed in Chapter 4.

A key part of the research was the **comparative modeling and interpretation phase**. A comparative experiment was designed and executed using five different modeling strategies to build a predictive model. Interpretability tools, such as SHAP, were used to deeply analyze the model's decision-making logic, addressing research questions RQ1 and RQ2. Finally, a systematic **policy application framework** was developed. This framework is designed to translate the model's technical outputs (including predictions and explanations) into actionable policy recommendations in the real world, thus answering research question RQ3.

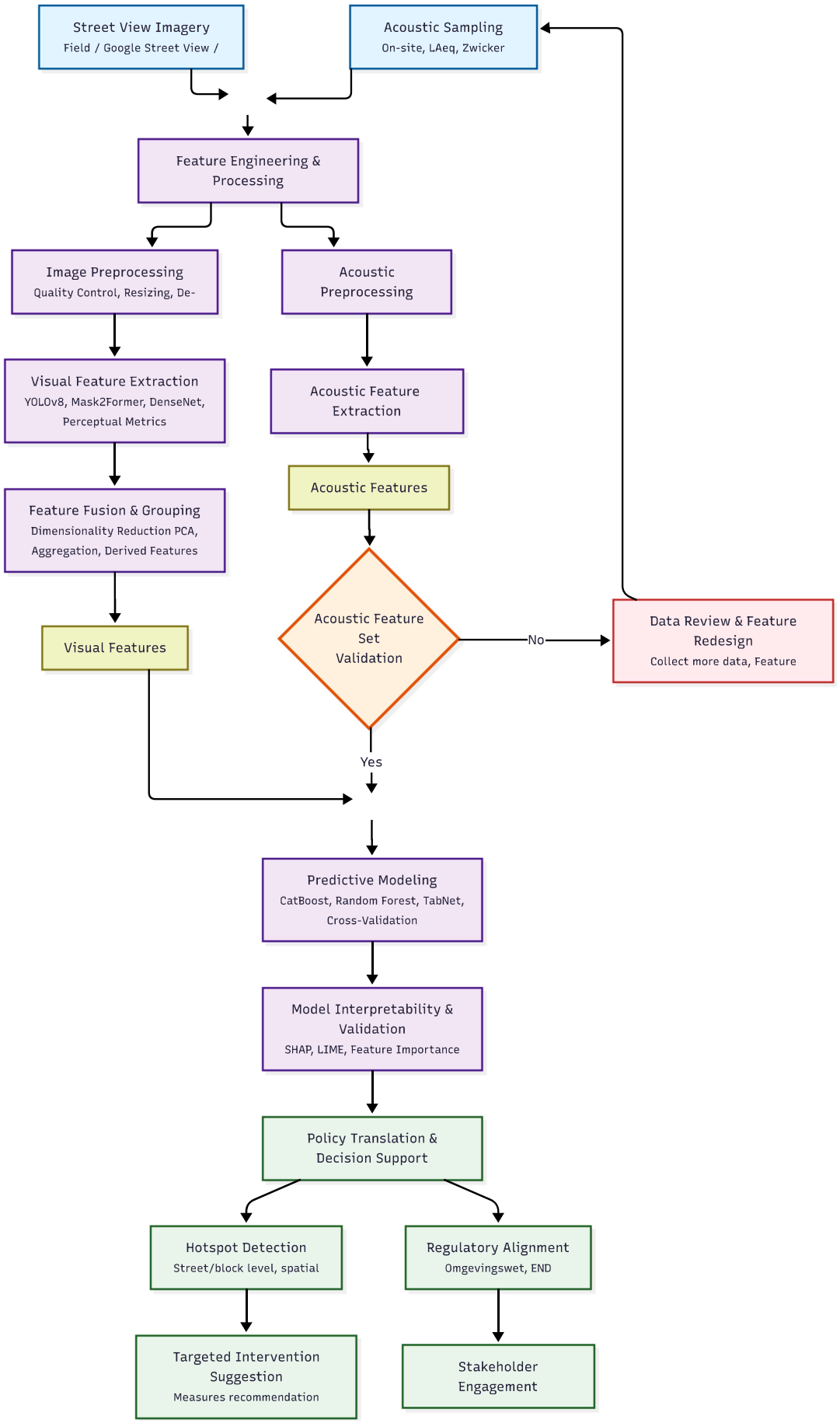


Figure 3.1: Urban Noise Flowchart

Figure 3.1 presents the end-to-end workflow of this research, from data acquisition to policy translation. The diagram uses a color-coded scheme to visually distinguish functional modules, with each color selected to convey intuitive associations in the research process.

- **Light blue blocks –Data Acquisition:** Represent the two primary input streams: Street View Imagery and Acoustic Sampling. Blue is commonly associated with reliability and objectivity, reflecting the factual, measurement-based nature of this stage. Detailed in Section 3.2 (3.2.1 and 3.2.2).
- **Light purple blocks –Technical Processing & Modeling:** Cover both the Feature Engineering & Processing chain—Image/Acoustic Preprocessing, Visual/Acoustic Feature Extraction, Feature Fusion & Grouping—and the Predictive Modeling & Interpretability stage. Purple, often linked to analytical thinking and synthesis, signals that these steps belong to the same computational and analytical "core stack." See Section 3.3 for feature engineering and Sections 3.5 and 5.3–5.4 for modeling and interpretability.
- **Light yellow (khaki) blocks –Intermediate Feature Sets:** Indicate the creation of structured data artifacts (Visual Features and Acoustic Features) that bridge raw inputs and modeling. A neutral color is used to suggest these are transitional outputs rather than final results. Described in Section 3.3 (visual features) and Sections 3.3–3.4 (acoustic features).
- **Orange diamond –Acoustic Feature Set Validation:** Marks a critical decision checkpoint where the suitability of acoustic features for prediction is assessed. Orange is chosen to draw attention to review and decision-making. Detailed in Section 3.4.
- **Pink/red blocks –Rework Loop:** Represent the corrective pathway (Data Review & Feature Redesign) triggered when the validation stage fails. The warm red tone signals caution and the need for corrective action, looping back to Sections 3.2–3.3.
- **Green blocks –Policy Translation & Decision Support:** Denote the final application stage, including Hotspot Detection (Section 6.3.1), Targeted Intervention Suggestion (Section 6.3.2), Regulatory Alignment (Section 6.3.3), and Stakeholder Engagement (Section 6.4.1). Green symbolizes forward movement and implementation. This stage fully corresponds to Chapter 6, where the decision chain—hotspot identification → causal diagnosis → measure selection → monitoring —is elaborated (Sections 6.2–6.4).

The color coding thus serves a dual role: enhancing diagram readability and reinforcing the conceptual grouping of research stages. Each color consistently maps to a specific phase of the methodology and its corresponding sections in the thesis, enabling readers to navigate between the visual workflow and the detailed technical descriptions.

3.2. Data Acquisition and Preparation

This study's data come from two complementary data streams. These streams are **street view images (SVI)**, which show the urban visual environment, and **on-site acoustic samples**, which are the acoustic environment's "ground truth."

3.2.1. Street View Image (SVI) Data Sources

This study used SVI data from three sources to capture diverse urban forms in the Netherlands and ensure complete data.

First, I collected data on-site with an Insta360 X4 panoramic camera. I rode or walked along planned routes in **Delft** and **The Hague**, Netherlands. This method ensured that each acoustic sample had a 360-degree visual scene from the exact same time and location, providing the highest quality paired data for model training. All collected videos were processed into standard image frames and resized to 224x224 pixels to meet the input requirements for later deep learning models.

Second, I downloaded additional image data from other major Dutch cities, such as **Amsterdam**, **Rotterdam**, and **Utrecht**, using public data platform APIs. This expanded the model's applicability to a wider area. I mainly used **Google Street View (GSV)**, selecting hundreds of representative geographic

sample points in each target city and downloading images from multiple angles (0°, 90°, 180°, 270°) to fully capture the 3D visual environment of the sample points.

I also used **Mapillary**, a crowdsourced platform, to fill gaps in GSV coverage, especially in areas like narrow streets, non-motorized vehicle lanes, or newly developed communities.

All images from public platforms were carefully checked. I used **image metadata** to verify timestamps and remove outdated images. More importantly, I used the **ClassifierQuality** and **ClassifierPanorama models** from **ZenSVI** (Ito et al., 2025), an open-source software library for SVI analysis, to automatically filter out blurry images, images with severe obstructions, or images with very poor lighting. I also ensured all images were panoramas, maintaining the quality and consistency of the data used in the model.

3.2.2. On-site Acoustic Data Collection

Acoustic data were collected simultaneously with on-site SVI images, creating an exact link between sight and sound. I used a **calibrated, sound level recorder** with a windscreen to reduce wind noise during movement, ensuring accurate acoustic measurements.

I extracted two main target variables from short audio clips, such as 10-second segments, at each sample point. The model predicts these variables. The first is **A-weighted equivalent continuous sound pressure level (**L_{Aeq}**)**, the most common physical measure used worldwide to assess environmental noise. It represents the average sound energy over time and is frequency-weighted based on human ear sensitivity. The second is **Zwicker Loudness**, a psychoacoustic metric based on the **ISO 532-1 standard**, measured in "sone." Unlike the purely physical L_{Aeq}, Zwicker loudness better models how loud humans perceive sounds, considering energy distribution at different frequencies and sound masking effects. Its formula is:

$$N = \int_0^{24} N'(z)\, dz \tag{3.1}$$

Here, $N'(z)$ is the loudness density in a specific Bark band, and z is the **Bark scale (0-24)**. The Bark scale is a psychoacoustic frequency scale proposed by **Eberhard Zwicker** (Zwicker, 1961), imitating how the human ear perceives frequency in a non-linear way. Predicting this metric allows my model to assess not just "how noisy" something is, but also "how loud" it sounds. To ensure accurate calculations, I used a device-specific calibration process to convert the recorded signals to physical sound pressure, then input this into the Zwicker model based on the **mosqito library** (Glesser et al., 2021).

Additionally, besides L_{Aeq} and loudness, which are my prediction targets, I also extracted over 20 detailed acoustic features from the audio, such as **Mel-frequency cepstral coefficients (MFCCs)**, which describe timbre, and **spectral centroid** and **spectral bandwidth**, which describe spectral shape. These features are not used as model inputs; instead, Chapter 4 uses them for in-depth exploratory analysis of soundscapes in different cities, confirming the study's main idea that different urban environments do have unique and machine-identifiable **"acoustic fingerprints."**

3.3. Feature Engineering: Quantifying the Urban Visual Environment

Feature engineering changes raw, unstructured image data into meaningful, structured numerical variables that machine learning models can use. This study created a complex and detailed feature engineering process. Its goal is to get as much rich visual information as possible from SVIs that relates to the acoustic environment.

3.3.1. Basic Visual Feature Extraction

A series of advanced computer vision models, which were pre-trained on large datasets, were used. From each street-view image (SVI), over 350 raw visual features were systematically extracted. These features covered several aspects, including specific objects, spatial layouts, and scene atmosphere.

For **Object Detection**, **YOLOv8 model (Yaseen, 2024)** was used. The YOLO (You Only Look Once) series of algorithms is known for balancing speed and accuracy. YOLOv8 is the latest version. It uses an advanced backbone network and feature pyramid network (FPN). This helps it efficiently find objects of many sizes in images. I used it to find and count over ten types of objects directly related to noise sources. These include **vehicles (car)**, **buses (bus)**, **pedestrians (person)**, and **bicycles (bicycle)**. These counts act as direct measures of traffic density and human activity.

For **Semantic Segmentation**, I used the **Mask2Former model (Cheng et al., 2022)**. This is an advanced, general-purpose image segmentation architecture. Its core is the **Swin Transformer**. Unlike traditional segmentation models, Mask2Former unifies segmentation tasks as "mask classification." It can precisely classify every pixel in an image. I used it to calculate the percentage of area taken up by different visual elements. I focused on elements like **buildings (building)**, **roads (road)**, **sidewalks (sidewalk)**, **sky (sky)**, and different types of **vegetation (vegetation)**. These features together show the physical makeup of urban spaces. This directly impacts how sound travels and reflects. The formula for calculation is:

$$Percentage_k = \frac{\text{Pixels of Class } k}{\text{Total Pixels}} \times 100 \tag{3.2}$$

To get high-level scene context, I used the **DenseNet161 deep neural network model** for **Scene Classification (Zhou et al., 2020)**. This model was pre-trained on the large **Places365 scene dataset**. It can predict the probability of each image belonging to one of 365 scene categories (like residential streets, highways, commercial areas). Its output is normalized using a **Softmax function**. This gives the probability p_i for each category:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{365} e^{z_j}} \tag{3.3}$$

Here, z_i is the network's raw output (logit value) for the i-th category.

The ZenSVI open-source library **ZenSVI (Ito et al., 2025)** was also used for SVI analysis. This enabled the quantification of several perceptual and technical metrics. The **perceptual scores**, such as **safety**, **liveliness**, and **beauty**,were predicted using a model trained on the **PlacePulse 2.0 dataset**. The technical metrics included **technical metrics** like **image clarity (quality)** and **lighting conditions (lighting)**.

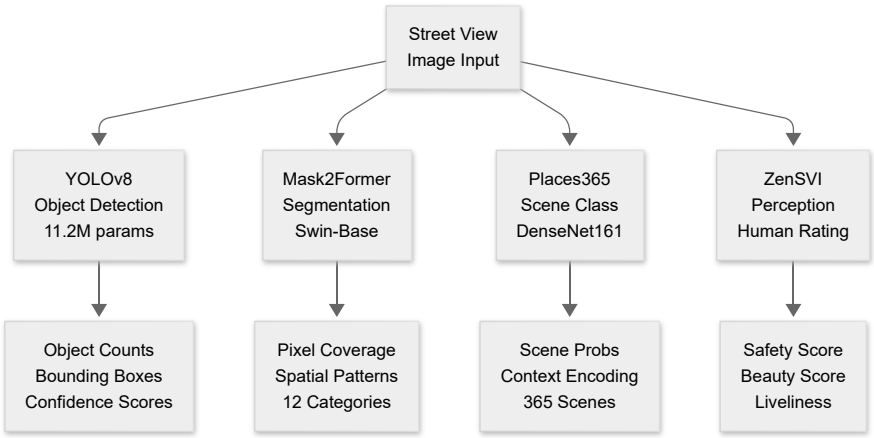


Figure 3.2: Multi-Model Pipeline Architecture

Finally, **Low-level Features** of the image were calculated. These include **global color distribution** (like the mean μ_c and standard deviation σ_c in the HSL color space) and **edge density**, calculated using the **Canny edge detector**. This captured the scene's overall tone, visual complexity, and texture information.

3.3.2. Advanced Feature Processing and Grouping

Having too many raw features can cause overfitting and slow down calculations if directly fed into the model. To address these issues, a fully automated, modular, and transparent feature engineering pipeline was implemented, allowing for flexible adaptation and rigorous quality control. This section details the methodology and mathematical foundations of the feature grouping, transformation, and optimization processes used in this study. The correspondence between these methods and their implementation in the code is also clarified.

1. Automated, Hierarchical Feature Grouping

The first stage of feature processing is the automated, semantic grouping of features. Instead of manual assignment, the pipeline uses a hierarchical, keyword-driven algorithm that parses each feature's name and assigns it to the most semantically relevant group. The assignment follows a strict priority. For example, features labeled with **scene_** are first checked against keywords related to transportation (e.g., "airport" or "station"), then against keywords for indoor, natural, and finally urban environments. Each feature is assigned to the first matching group, ensuring mutual exclusivity and domain consistency.

Formally, this process can be described by a mapping function $g : F \to G$, where F is the set of all features and G is the set of predefined groups. For each feature $f \in F$, the group assignment is determined by the following equation:

$$g(f) = \underset{g \in G}{\operatorname{argmin}} \left[\mathbb{I}(\text{feature } f \text{ matches group } g \text{ keywords}) \right] \tag{3.4}$$

Here, $\mathbb{I}(\cdot)$ is the indicator function, which returns 1 if the feature matches the group's keyword set and 0 otherwise. This ensures that each feature is assigned to at most one group, making the assignment reproducible and interpretable. In the code, this logic is implemented in the `UpdatedDataManager` class by the `_update_feature_groups` method, which supports dynamic adjustments whenever the feature set changes.

2. Group-Specific Feature Engineering

Within each group, the pipeline applies customized transformations to enhance the informativeness and comparability of features. For **color features**, the diversity of the color distribution is quantified using **Shannon entropy**:

$$H = -\sum_{i=1}^{N} p_i \log(p_i + \epsilon) \tag{3.5}$$

Here, p_i is the proportion of pixels assigned to the i-th color bin, N is the total number of color bins, and ϵ is a small constant to prevent numerical instability. This entropy value reflects the visual complexity of an image: high entropy indicates a visually rich scene, whereas low entropy indicates a scene dominated by a single color. In addition, other metrics like color contrast and a dominant color index are also generated.

For **object count features**, a **logarithmic transformation** is applied:

$$x' = \log(1 + x) \tag{3.6}$$

Here, x is the raw count and x' is the transformed value, which reduces the impact of extreme values and normalizes the scale across categories. The diversity of object types is captured by counting the number of nonzero categories:

$$\text{Diversity} = \sum_{j=1}^{M} \mathbb{I}(x_j > 0) \tag{3.7}$$

Here, x_j is the count for the j-th object or segment category. This metric provides a simple measure of categorical richness within a scene.

For **categorical probability features**, such as scene or segmentation outputs, the pipeline extracts the **dominant class index** and its associated **confidence**:

$$\text{Dominant Index} = \underset{j}{\operatorname{argmax}}(x_j) \tag{3.8}$$

$$\text{Confidence} = \max_j(x_j) \tag{3.9}$$

Here, x_j is the predicted probability or area proportion for category j. These features work together to describe both the most likely semantic class and the certainty of the prediction.

All these group-specific engineering steps are implemented in a modular way within the data manager, ensuring that any new or re-defined group can automatically benefit from the appropriate transformations.

3. Dimensionality Reduction via PCA

Many feature groups, such as scene probabilities or fine-grained segmentation maps, are inherently high-dimensional. To avoid redundancy and overfitting, **Principal Component Analysis (PCA)** is applied within each such group to extract the most informative linear combinations of features. The PCA process consists of four canonical steps.

Mean Centering Each feature is centered by subtracting its mean:

$$X'_{ij} = X_{ij} - \bar{X}_j \tag{3.10}$$

Here, X_{ij} is the raw value, and $\bar{X}_j$ is the mean of feature j.

Covariance Matrix Computation The covariance matrix Σ is computed as:

$$\Sigma = \frac{1}{n-1}(X')^\top X' \tag{3.11}$$

Here, n is the number of samples.

Eigen-Decomposition The principal axes are obtained by solving the equation:

$$\Sigma v = \lambda v \tag{3.12}$$

This process yields **eigenvalues (λ)** and **eigenvectors (v)**.

Component Selection and Projection The top k principal components (eigenvectors with the largest eigenvalues) that explain a desired cumulative variance (e.g., 90%) are selected, and the data is projected as follows:

$$Y = X'W \tag{3.13}$$

Here, W is the matrix of selected eigenvectors, and Y is the reduced representation.

The number of components and the variance threshold are dynamically configurable in the pipeline. This step reduces computational cost and enhances the signal-to-noise ratio, preserving the essential structure of each group. In the code, this is implemented in the `_process_scene_features`, `_process_segmentation_fea` and similar methods, using the `sklearn.decomposition.PCA` class.

4. Composite Perceptual Indices

For perceptual evaluation features, the pipeline computes mean scores for positive and negative attributes to derive an overall **polarity score**:

$$\text{Polarity} = \frac{1}{N^+}\sum_{i=1}^{N^+} x_i^+ - \frac{1}{N^-}\sum_{j=1}^{N^-} x_j^- \tag{3.14}$$

Here, x_i^+ and x_j^- are the positive and negative attribute scores, while N^+ and N^- are their respective counts. This scalar value summarizes the overall affective impression of a scene, with higher values indicating a more favorable perception. Optionally, PCA can be applied to the full set of perceptual attributes to capture latent factors, which can then be used for dimensionality reduction or further modeling.

5. Automated Group Quality Control and Optimization

This links to the "Feature fusion & Grouping" in Figure 3.1. A major innovation of this work is the introduction of a closed-loop, automated quality control system for feature groups. After initial feature engineering, the pipeline uses a comprehensive validator module to assess each group based on several quantitative metrics: **average intra-group correlation** (consistency), **proportion of highly correlated pairs** (redundancy), **feature sparsity**, and **group size**.

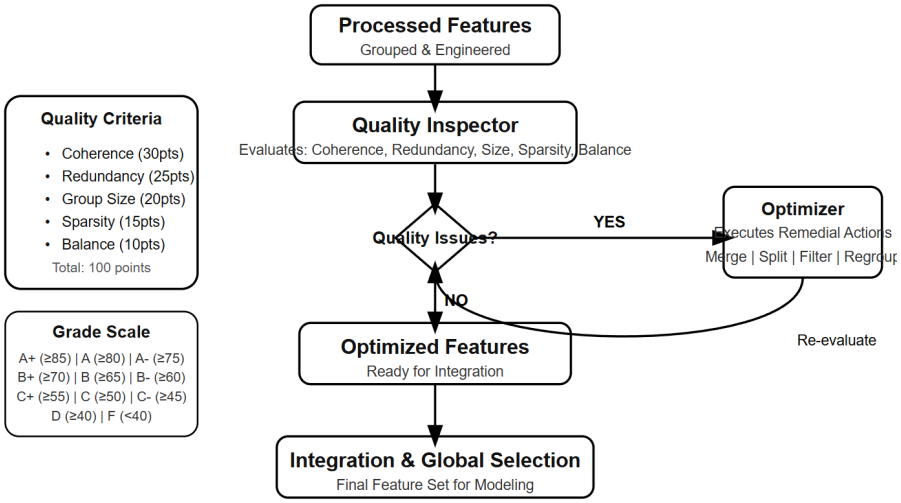


Figure 3.3: Quality Control & Iterative Optimization Workflow

These metrics are integrated into a weighted score:

$$Q_{\text{score}} = \sum_{i=1}^{N} w_i \cdot \text{Metric}_i(\text{group}) \tag{3.15}$$

Here, w_i is the weight for metric i, and Metric_i is the corresponding quantitative function. The Q_{score} is mapped to a grade (from A+ to F), providing a data-driven basis for further action.

For groups with suboptimal grades, automated optimization strategies are triggered. For instance, scene groups with low internal consistency are merged and re-clustered using hierarchical clustering; very large or semantically mixed groups (e.g., daily life objects) are split based on domain-driven subcategories; and groups with excessive redundancy undergo correlation-based feature filtering. After each optimization, group quality is automatically reassessed, forming a closed loop.

This entire process is fully programmatic: every step, from validation to optimization and revalidation, is orchestrated by modular, reusable code (`UpdatedGroupingValidator`, `UpdatedFeatureGroupOptimizer`) and can be invoked iteratively as needed. The result is a feature set that is both information-rich and structurally robust, laying a solid foundation for high-performance and interpretable predictive modeling.

3.4. Methodology for Acoustic Feature Set Validation

Before visual features were used to predict the acoustic environment, a logical point was checked. The question was whether the acoustic feature set, which served as the prediction target, contained enough stable and useful information to differentiate between various urban environments. This was to determine if the ”soundscapes” of different cities have unique ”acoustic fingerprints” that could be identified by a machine. If this were not the case, any further predictive modeling would lack a solid foundation.

To systematically address this question, I designed a complete method. It includes **statistical testing**, **structure analysis**, and **machine learning classification validation**. These methods are to be used in Chapter 4 to assess the quality and distinctiveness of the extracted acoustic feature set.

3.4.1. Statistical Difference Analysis

To check if the acoustic features collected from different places (like **Delft** and **The Hague**) show significant differences in their distribution, two statistical analysis methods were used.

First, the **Mann-Whitney U non-parametric test**was employed. This statistical test is used to check if two independent samples have significantly different distributions. This method was chosen because, unlike parametric tests like the t-test, it does not require the data to follow a normal distribution or for the two sample groups to have equal variance, making it suitable for acoustic feature data that may not be normally distributed. Its null hypothesis (H_0) is that the two independent samples' distributions are not significantly different. he null hypothesis would be rejected if the calculated p-value was less than the chosen significance level (e.g., 0.05). This would indicate that the distributions of acoustic features from the two locations are statistically different.

However, a p-value only tells us if a difference is “significant.” It does not measure the “size” or practical importance of the difference. So, **Cohen's d effect size** was used as a supplement. Cohen's d is a standardized effect size measure. It quantifies the size of the difference between two group means. Its formula is:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_p} \tag{3.16}$$

Here, $\bar{x}_1$ and $\bar{x}_2$ are the means of the two samples. s_p is the pooled standard deviation. Its calculation is $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$. Generally, $d \approx 0.2$ is considered a small effect, $d \approx 0.5$ a medium effect, and $d \approx 0.8$ a large effect. By calculating Cohen's d, the effectiveness of different acoustic features in distinguishing urban environments can be better understood.

3.4.2. Internal Structure Analysis of Acoustic Features

In addition to checking for differences, the internal structure and relationships within the acoustic feature set needed to be understood. For this purpose, **Principal Component Analysis (PCA)**, as described in detail in Section 3.3.2, was used. PCA was applied to the complete 21-dimensional acoustic feature set with two goals: 1) to explore whether these features could be grouped into a few unrelated, combined dimensions that represent core acoustic concepts (such as ”loudness” or ”timbre brightness”); and 2) to visualize the data and see if acoustic samples from different cities form distinct clusters in the reduced principal component space.

3.4.3. Machine Learning-Based Soundscape Discriminability Validation

The most comprehensive test of ”acoustic fingerprint” effectiveness is to determine if a machine learning model can accurately classify the source location using only acoustic features.

For this purpose, a **binary classification task** was designed. The 21 acoustic features were used as the sole input to train a machine learning classifier to determine whether an audio clip originated from **Delft** or **The Hague**. The performance of several classification algorithms was evaluated, with a focus on the **Support Vector Machine (SVM)**. SVM is a powerful supervised learning model that works by finding an **optimal hyperplane** in a high-dimensional space to separate different classes of samples

with the largest possible margin. SVM was chosen because it typically performs well and efficiently for classification problems with small to medium-sized, high-dimensional datasets. If the model achieves high accuracy, significantly better than random guessing (50%), it would strongly support the hypothesis that ”different urban soundscapes have unique characteristics that machines can identify.”

3.5. Predictive Modeling and Interpretability Framework (Answering RQ1 & RQ2)

This section describes the modeling and explanation framework designed to answer research questions **RQ1 (identifying key visual features)** and **RQ2 (building explainable predictive models)**. The main idea is to use a **comparative experimental strategy** to systematically evaluate the effectiveness of different modeling approaches. Strong interpretability tools are also included to analyze how the best model makes decisions.

3.5.1. Comparative Modeling Strategies and Algorithm Principles

To fully and objectively assess if visual features can predict acoustic features, and to avoid bias from relying on just one model, I designed and ran **five parallel model pipelines**. Each pipeline represents a different modeling philosophy and technical approach. Comparing their results will help provide a deeper understanding of the problem's complexity and the suitability of different methods.

Pipeline 1: Optimal Pipeline - CatBoost & Random Forest

This is the main method recommended for this study. It uses the final feature set, which was fully engineered as described in Section 3.3, including **PCA dimensionality reduction** for certain high-dimensional feature groups. Based on preliminary experiments and a review of the literature, two powerful **ensemble learning models** were selected and optimized separately for two different prediction goals.

CatBoost (Categorical Boosting) was chosen as the main model for predicting the physical sound pressure level, L_{Aeq}. CatBoost is an advanced algorithm based on **Gradient Boosting Decision Trees (GBDT)**. It addresses common GBDT issues with two key innovations. First, it uses an **Ordered Boosting strategy** to handle **target leakage**, which is particularly important for categorical features. Second, it uses **symmetric trees (oblivious trees)** as base learners, meaning all nodes at the same tree level use the same feature for splitting. This makes the model's predictions faster and also acts as a regularizer, preventing the model from becoming overly complex. Its objective function is similar to traditional GBDT. It aims to minimize the sum of the loss function $\mathcal{L}$ and the regularization term Ω:

$$\text{Objective} = \sum_{i=1}^{n} \mathcal{L}(y_i, F(x_i)) + \sum_{t=1}^{T} \Omega(f_t) \tag{3.17}$$

Here, $F(x_i)$ is the model's predicted value for sample x_i. It's built by adding up multiple trees f_t.

Random Forest (RF) was chosen as the main model for predicting the psychoacoustic loudness, **Zwicker Loudness**. RF is an **ensemble learning algorithm** based on the idea of **Bagging (Bootstrap Aggregating)**. It improves model stability and accuracy by building many independent decision trees. Its building process has two main ”random” steps: 1) **Bootstrap** random sampling with replacement of training samples to create a different training subset for each tree; 2) when splitting each tree node, it chooses from a randomly selected subset of features, instead of searching for the best split point from all features. These two random processes ensure that each tree is different, which effectively reduces the model's variance. For regression tasks, the final prediction is the average of the predicted values from all T decision trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \tag{3.18}$$

Here, $h_t(x)$ is the prediction of the t-th tree.

Pipeline 2: Baseline Pipeline - GBRT

This pipeline serves as a strong baseline to determine how much **advanced feature processing (specifically PCA dimensionality reduction)** contributes to performance. It uses a feature set that has undergone only basic derived feature calculations, without PCA dimensionality reduction. A standard **Gradient Boosting Regressor (GBRT)** model is used for prediction. GBRT is an additive model that iteratively trains new decision trees to fit the residuals from the previous iteration. In step m, the model $F_m(x)$ is updated as:

$$F_m(x) = F_{m-1}(x) + \nu \cdot h_m(x) \tag{3.19}$$

Here, $F_{m-1}(x)$ is the ensemble model of the first $m-1$ trees, ν is the **learning rate**, and $h_m(x)$ is the new decision tree trained to fit the **negative gradient** (which is the residual) of the current loss function concerning $F_{m-1}(x)$.

Pipeline 3: Deep Learning Pipeline - Multi-Branch Network

This pipeline explores the potential of deep learning for handling this type of mixed tabular data. I designed and built a custom **Multi-Branch Neural Network**. This network has a separate, shallower neural network branch for each semantic feature group (like ”color,” ”vehicles,” ”natural environment segmentation”). Each branch has multiple **fully connected layers (nn.Linear)**, **batch normalization (nn.BatchNorm1d)**, and **ReLU activation functions**. These learn a low-dimensional abstract representation of that feature group. Then, the outputs of all branches are **concatenated** and sent to a deeper fusion network for the final combined prediction.

Pipelines 4 and 5: Baseline Pipelines - Ridge & Plain Linear

These two pipelines represent the most basic modeling methods. They serve as reference points for all the more complex models. Both are trained on the original feature set without PCA dimensionality reduction.

Ridge Regression is an improved least squares estimation method. It solves the instability of ordinary linear regression when dealing with multicollinear features by adding an **L2 regularization term** to the loss function:

$$\text{Objective} = \sum_{i=1}^{n}(y_i - w^T x_i)^2 + \alpha \sum_{j=1}^{p} w_j^2 \tag{3.20}$$

Here, w is the model's weight vector, and α is a hyperparameter that controls the strength of regularization.

Plain Linear Regression finds a set of weights w to minimize the **Residual Sum of Squares (RSS)** between predicted and actual values.

3.5.2. Model Training, Optimization, and Validation

To ensure all models performed optimally and were evaluated fairly and reproducibly, a consistent and strict training and validation process was used.

First, in the **data splitting phase**, machine learning best practices were strictly followed. The complete dataset was divided into an **80:20 ratio** for the **training set** and a permanently held-out **test set**. The training set was then further split at a **75:25 ratio** into a **sub-training set** (for model learning) and a **validation set** (for hyperparameter tuning). The entire splitting process used a **fixed random seed (random_state=3407) (Picard, 2021)** to ensure reproducibility.

Second, in the **optimal model pipeline**, **feature selection** was performed. A **Random Forest model**, trained on the training set, was used to evaluate the importance of all features. The **top 80 core features** that contributed most to the prediction target were automatically selected using `sklearn.feature_selection.Sele`
This reduced dimensionality and noise and lowered the risk of overfitting.

Next, during the **data scaling phase**, features were standardized according to model requirements. For the optimal pipeline, **RobustScaler** was used, which subtracts the median and divides by the **Interquartile Range (IQR)**, making it robust to outliers. Other pipelines used **StandardScaler**. All scalers were `fit` on the training set and the `transform` was applied to the validation and test sets to prevent **data leakage**.

In the **hyperparameter optimization phase**, **Optuna**, a **Bayesian optimization framework**, was used. For each core model (**CatBoost, RF, GBRT, Ridge**), hyperparameters were searched using the **Tree-structured Parzen Estimator (TPE)**. The performance of each parameter set was evaluated using **K-Fold Cross-Validation (K=5)**. This method splits the training set into K subsets, trains on K-1, and validates on the remaining one. The process is repeated K times, and the results are averaged to produce a more robust performance estimate.

Final model performance was measured using standard metrics:

- **Root Mean Squared Error (RMSE):** $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
- **Mean Absolute Error (MAE):** $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
- **Coefficient of Determination (R^2):** $R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$
- **Mean Absolute Percentage Error (MAPE):** a relative error measure

3.5.3. Model Interpretability Framework

Answering **RQ2 (how to build an explainable model)** was a top priority. To prevent machine learning models from becoming **"black boxes"** and to increase confidence in policy contexts, two **model-agnostic interpretability methods** were used.

SHAP (SHapley Additive exPlanations)

SHAP is a unified framework for interpreting machine learning model predictions, grounded in **Shapley values** from cooperative game theory. The Shapley value provides a theoretically fair way to attribute the contribution of each feature to a model's prediction by considering all possible feature combinations.

For a model f and input x, the SHAP value ϕ_j for feature j is defined as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(M - |S| - 1)!}{M!} \left[f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S) \right] \tag{3.21}$$

where F is the set of all features, M is the total number of features, S is a subset of features not containing j, and $f_S(x_S)$ is the model prediction using only features in S. This formula averages the marginal contribution of feature j over all possible feature orderings.

The final prediction can be decomposed as:

$$f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j \tag{3.22}$$

where ϕ_0 is the expected model output (base value) and ϕ_j is the SHAP value for feature j.

SHAP provides both **global** and **local** interpretability:

- **Global analysis:** Aggregates SHAP values across the dataset to produce feature importance rankings and summary plots, revealing which features most influence predictions overall (answers RQ1).
- **Local analysis:** For a specific instance, SHAP force plots visualize how each feature pushes the prediction higher or lower compared to the base value, explaining individual decisions (answers RQ2).

- **Directionality and magnitude:** SHAP summary dot plots show both the direction (positive/negative impact) and the magnitude of each feature's effect.

In this study, SHAP is used to systematically analyze both the overall decision logic of the model and the specific factors driving predictions for key samples (such as noise hotspots), ensuring transparent and trustworthy model interpretation.

LIME (Local Interpretable Model-agnostic Explanations)
As a complement to SHAP, **LIME** (Local Interpretable Model-agnostic Explanations) was used to further enhance model interpretability. LIME works by locally approximating the complex, potentially non-linear prediction function $f(x)$ with a simple, interpretable surrogate model $g(x)$ (typically linear or decision tree) around the instance of interest x_0. The process involves generating a set of perturbed samples $\{x_i'\}$ near x_0, obtaining their predictions $f(x_i')$, and then fitting $g(x)$ to these samples using a weighted loss function that emphasizes proximity to x_0.

The objective function for LIME is:

$$\arg\min_{g \in G} \mathcal{L}(f, g, \pi_{x_0}) + \Omega(g) \tag{3.23}$$

where $\mathcal{L}$ is a loss function measuring the fidelity of g to f in the neighborhood defined by the proximity kernel π_{x_0}, and $\Omega(g)$ is a complexity penalty to encourage interpretability.

For each feature j, the local linear surrogate assigns a weight w_j indicating its contribution to the prediction at x_0:

$$g(x) = w_0 + \sum_{j=1}^{M} w_j x_j \tag{3.24}$$

LIME was applied to explain the predictions of selected extreme or outlier samples, providing intuitive, locally faithful explanations. The results were compared with SHAP's local interpretations to verify consistency and robustness. This dual approach ensures that both global and local model behaviors are transparent, supporting reliable decision-making in policy contexts.

3.6. Implementation Translation Framework (Answering RQ3)

To systematically answer research question **RQ3** on how to translate the model's technical output into practical implementation management of technology recommendations, this study designed a rigorous and multi-layered **policy translation framework**. This framework aims to build a solid bridge from complex data insights to practical, actionable measures, ensuring that the scientific findings of this study can support urban environmental management and planning decisions in the most efficient and powerful way possible.

3.6.1. Spatial Hotspot Identification

The first step in policy implementation is to pinpoint the problem areas. In this study, a trained and optimized machine learning model was used to perform large-scale batch predictions for all densely sampled locations within the target area. This generated high-resolution geospatial distribution maps of the A-weighted equivalent continuous sound pressure level (L_{Aeq}) and Zwicker loudness with street-level accuracy. These detailed maps can reveal localized high-noise "hotspots" that are often overlooked by traditional, coarse-grained noise assessment methods.

Hotspots were defined using two complementary practical standards:

- **Absolute Threshold Method**: Based on World Health Organization (WHO) health risk recommendations, locations where the model's predicted values frequently or consistently exceed a specific health risk threshold (e.g., $L_{\mathrm{Aeq}} > 65$ dBA) were identified as "health risk hotspots." This provides justification for direct intervention by public health authorities.
- **Relative Ranking Method**: This method identified the areas with the highest relative noise levels within the city (e.g., the top 5% or 10% of areas by noise level) and defined them as "relative

noise hotspots." This approach is crucial for fine-grained urban management, allowing cities to prioritize areas that have a significant impact on residents' quality of life based on their specific characteristics and resources.

3.6.2. Causal Diagnosis Process Based on Interpretability

After identifying hotspots, the second step of the framework is to use the interpretability tools introduced in Section 3.5.3 (SHAP) to systematically diagnose the causes of the noise hotspots identified in the first step. The purpose of this method is to link abstract noise levels to specific, actionable urban visual elements. The diagnostic process involves analyzing the SHAP values for samples in each hotspot area to identify which visual features (such as green space ratio, building density, or scene perception features) are the key drivers of the high noise predictions. This step provides direct, data-driven clues for formulating "symptomatic" intervention measures.

3.6.3. Data Foundation for Dual-Metric Evaluation

A key part of this research methodology is the development of a dual-target model capable of predicting both physical noise (L_{Aeq}) and perceived loudness (Zwicker Loudness) simultaneously. By providing predictions for both of these metrics for the same geographical location, the method establishes the necessary data foundation for a multi-dimensional soundscape quality assessment. This allows subsequent policy analysis to go beyond single-metric physical compliance and examine areas that may be "physically compliant but perceptually noisy," thereby providing a basis for more human-centered soundscape planning decisions.

3.6.4. Regulatory Alignment and Decision Support Protocol

To ensure the practical applicability of the research findings, this framework includes a set of regulatory alignment protocols. The core method of this protocol is to use standardized correction factors based on typical diurnal noise curves to approximate the model's short-term L_{Aeq} metric into the long-term metrics required by regulations such as the European Union Environmental Noise Directive (END) (e.g., L_{den}). At the same time, this protocol requires that all model outputs must be accompanied by a quantification of their uncertainty (such as RMSE) to ensure that decision-makers fully understand their accuracy limitations, thereby supporting a robust decision-making process.

3.6.5. Dissemination and Adoption

To guide the model's transition from a research prototype to a practical application, a strategic adoption roadmap framework was developed. This framework uses the Gartner Hype Cycle (Gartner, Inc., 2025) (see Figure 3.4), which typically describes the evolutionary path of new technologies from the "Innovation Trigger" to the "Plateau of Productivity." Fully recognizing the "Peak of Inflated Expectations" and "Trough of Disillusionment" that can occur during the adoption of new technologies is crucial for managing the introduction of AI noise models to stakeholders. This study advocates a **phased, iterative adoption strategy** aimed at gradually building validation and trust to avoid common pitfalls caused by over-promotion or unrealistic expectations.

This methodology divides the model's widespread adoption into four logical phases: 1) **pilot validation**, which involves small-scale applications in a few innovative cities to prove the concept; 2) **capacity building and knowledge sharing**, which disseminates successful cases and lessons learned through professional networks (e.g., VNG); 3) **platform and infrastructure development**, which involves creating tools or cloud platforms for large-scale use; and 4) **institutional integration**, which formally incorporates the method into standard city planning and environmental assessment procedures. This framework provides methodological guidance on how to manage technological expectations, build trust, and ultimately seamlessly integrate this tool into existing workflows.

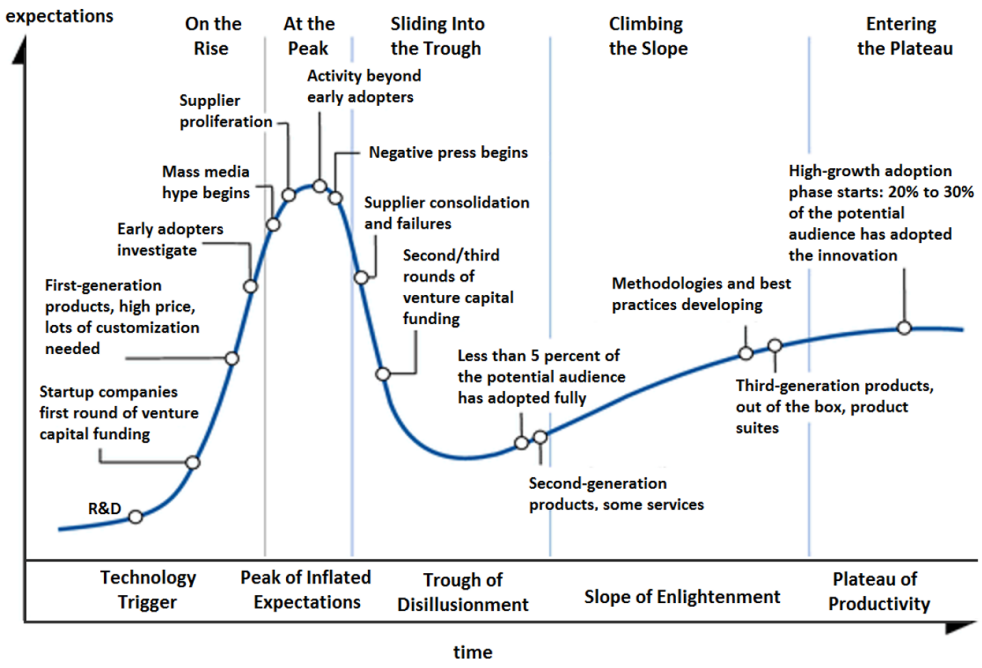


Figure 3.4: Gartner Hype Cycle

4

Acoustic Data Exploration and Feature Analysis

4.1. Introduction

Before the complex visual features were used to predict the urban acoustic environment in Chapter 5, this chapter undertook the crucial task of establishing the **scientific legitimacy** of the entire study. The research first needed to scientifically prove that the set of acoustic features targeted for prediction was not random noise, but rather contained stable, distinct, and internally structured geospatial information. If ”acoustic fingerprints” did not exist, all subsequent predictive modeling would lack a logical foundation.

Therefore, the objective of this chapter was not to build a predictive model directly, but to conduct an in-depth exploratory analysis and validation of the acoustic data itself, which served as the ”ground truth.” The 21-dimensional acoustic feature set described in Chapter 3 was systematically examined to ultimately answer a fundamental question: Do different urban environments truly possess unique, machine-recognizable ”acoustic fingerprints”?

To answer this question, a progressive analytical framework was adopted in this chapter. First, nonparametric statistical tests and effect size analysis were used to statistically determine whether significant differences existed in the acoustic features of different cities, represented by Delft and The Hague. Next, correlation analysis and principal component analysis (PCA) were employed to explore the internal relationships and potential core dimensions of these acoustic features. Finally, and most critically, a machine learning classification task was designed to test whether an objective algorithmic model (a support vector machine) could accurately classify the source city of an audio clip based solely on its acoustic features.

The results of the analysis in this chapter will lay a solid data and theoretical foundation for the rest of the study. If it can be successfully demonstrated that urban soundscapes possess stable and machine-recognizable ”acoustic fingerprints,” then the cross-modal prediction task from visual to auditory data in Chapter 5 will not only be logically sound but also strongly supported by the data.

4.2. Analysis of Acoustic Feature Differences Between Locations

4.2.1. Statistical Validation of Feature Differences

The first step of the analysis was to directly compare the distribution of all 21 acoustic features between the two cities. I conducted an independent samples Mann-Whitney U non-parametric test for all features to determine the statistical significance of their mean differences, and I also calculated Cohen’s d to measure the effect size. This analysis led to a core and powerful finding: Across all five feature groups covered in this study, **all 21 acoustic features showed extremely significant statistical differences between Delft and The Hague ($p < 0.05$)**. This finding provides a solid statistical foundation for my central premise that acoustic features contain location-specific information.

A deeper look at the specific statistical data reveals detailed patterns of difference. First, regarding the physical and psychoacoustic metrics that describe the fundamental energy and perception of the sound environment, The Hague's mean $L_{eq,\text{dBA}}$ (A-weighted equivalent sound pressure level) and Zwicker loudness (sone) were significantly higher than Delft's. This aligns with the general understanding that larger cities typically have higher background noise levels. Similarly, for spectral and temporal features describing sound frequency composition, The Hague's metrics, such as **spectral centroid mean** and **zero-crossing rate mean**, were also significantly higher. This suggests that The Hague's soundscape may contain more high-frequency components or more frequent transient events.

However, the most revealing differences were found in the **Mel-Frequency Cepstral Coefficients (MFCCs)**. All 13 MFCCs showed significant differences, with several coefficients exhibiting very large effect sizes. For example, the Cohen's d value for **mfcc_13_mean** was an astonishing 2.09, while **mfcc_8_mean** and **mfcc_9_mean** also reached 1.26 and 1.23, respectively. Such large effect sizes indicate that these mid-to-high order MFCCs, which are typically related to the fine spectral structure and timbre of sounds, are extremely sensitive indicators for distinguishing the soundscapes of the two cities. They capture more subtle "texture" differences in the soundscape, going beyond simple loudness or energy. Figure 4.1 visually presents these differences through box plots.[1]

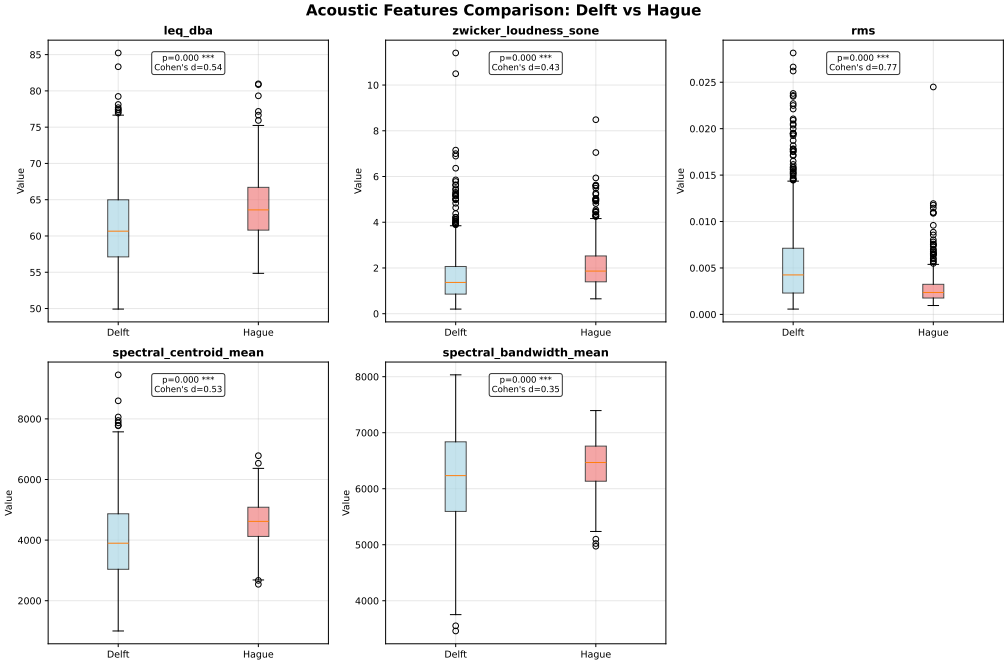


Figure 4.1: Distribution Comparison of Feature Groups in Delft and The Hague Box Plots

4.2.2. Distribution Patterns of MFCC Features

To better understand the **MFCCs**, which form the core "fingerprint" of the soundscape, I examined their probability density distributions, as shown in Figure 4.2 [2]. These distribution plots not only confirmed the statistical differences in means but also revealed the distinct "personalities" of the soundscapes in both cities.

[1] Figure Description:
(a) Box Plot Comparison of Selected Physical Acoustic and Spectral Features: This graph shows box plots for some of the physical acoustic and spectral features, comparing their distributions between Delft and The Hague. All p-values are less than 0.001, indicating extremely significant differences between the two cities for these features.
(b) Radar Chart Comparison of Five Feature Groups: This radar chart presents a normalized comparison of all five feature groups. The normalization allows for cross-metric comparison, clearly illustrating the distinct "soundscape profiles" of the two cities across various feature dimensions.

[2] Each subplot displays the **probability density distribution** for a single MFCC coefficient. The asterisks ($*$, $**$, $***$) in the plots indicate **statistical significance**. The **blue dashed line** represents the mean position for Delft, while the **red dashed line** indicates the mean position for The Hague. Darker Areas: These indicate the overlap between the two distributions, where the MFCC features from Delft and The Hague coincide.

For example, looking at **mfcc_5_mean** and **mfcc_8_mean**, you can see that the distributions for the two cities are almost completely separate, with their peak positions far apart. For other features, like **mfcc_2_mean**, even though the effect size of the mean difference isn't large ($d = 0.088$), there are still visible differences in the shape and dispersion of the distributions. This visual evidence further confirms that MFCCs provide unique, quantifiable signatures for the soundscapes of the two cities, not just in their numerical values but also in their overall statistical distribution patterns.

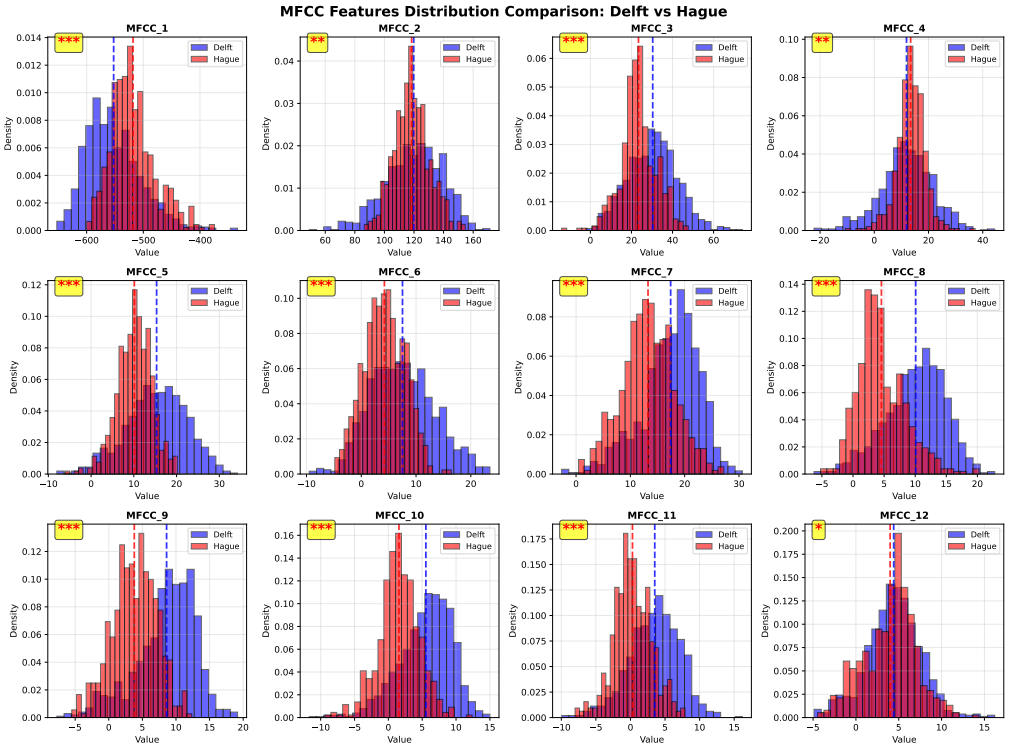


Figure 4.2: Probability Density Distribution Comparison of MFCC Coefficients in Delft and The Hague

4.3. Internal Structure and Correlation of Acoustic Features

4.3.1. Analysis of Feature Correlations and Structure

After confirming the discriminative power of individual features, understanding their internal relationships is crucial. This helps us identify redundant features and provides insights for subsequent model building. For this study, I performed a **correlation analysis**, which showed that 22 pairs of features had a strong correlation (Pearson correlation coefficient $|r| > 0.7$). This indicates a complex internal structure within the feature set. Figure 4.3, 4.4 visually present these intricate dependencies through a heatmap and a network graph.

These correlation analyses reveal several important structural characteristics. First, there's significant **high correlation within groups**, especially prominent in the **Spectral feature group**. For example, **spectral_centroid_mean** and **spectral_rolloff_mean** have a very high correlation coefficient of 0.973. This suggests they largely capture similar spectral information, which could lead to **multicollinearity issues** when building models, requiring careful handling.

Second, I observed widespread **high correlation across different groups**. A typical example is $L_{eq,\text{dBA}}$ (physical acoustic), which is highly correlated not only with **zwicker_loudness_sone** (psychoacoustic, $r = 0.945$) from the same group but also with **mfcc_1_mean** ($r = 0.914$) from the MFCC group. This tight coupling reveals that a soundscape's objective energy, subjective loudness perception, and basic spectral envelope are intrinsically unified and interconnected.

Finally, the analysis also found some **strong negative correlations**. For instance, **mfcc_2_mean** shows a strong negative correlation with multiple spectral features, including **spectral_centroid_mean**

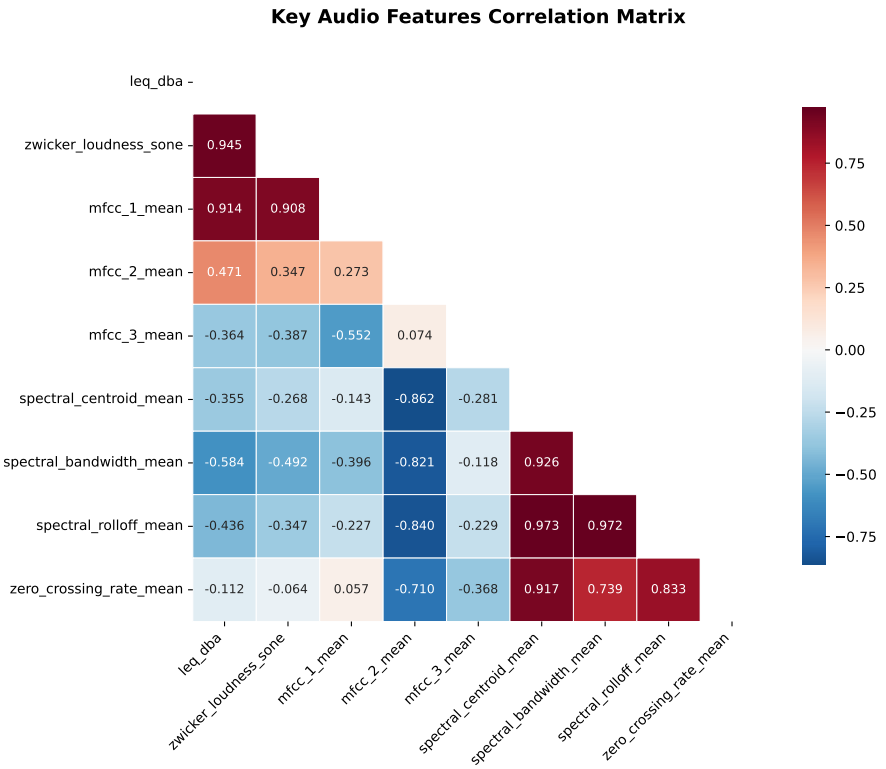


Figure 4.3: Correlation Structure of Audio Features Heatmap

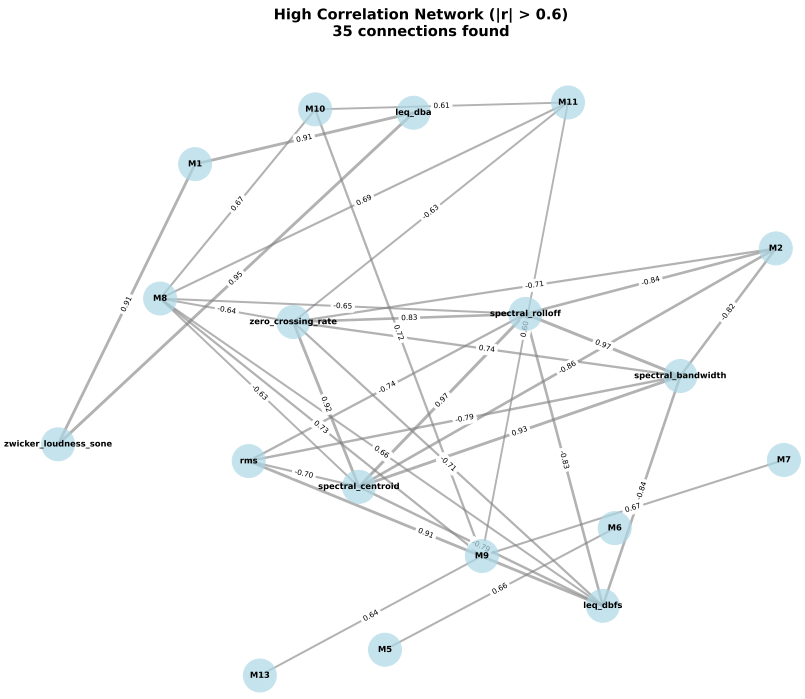


Figure 4.4: Correlation Structure of Audio Features Network Graph

($r = -0.862$). This might suggest some kind of physical or perceptual antagonism and trade-off between different dimensions of the soundscape features.

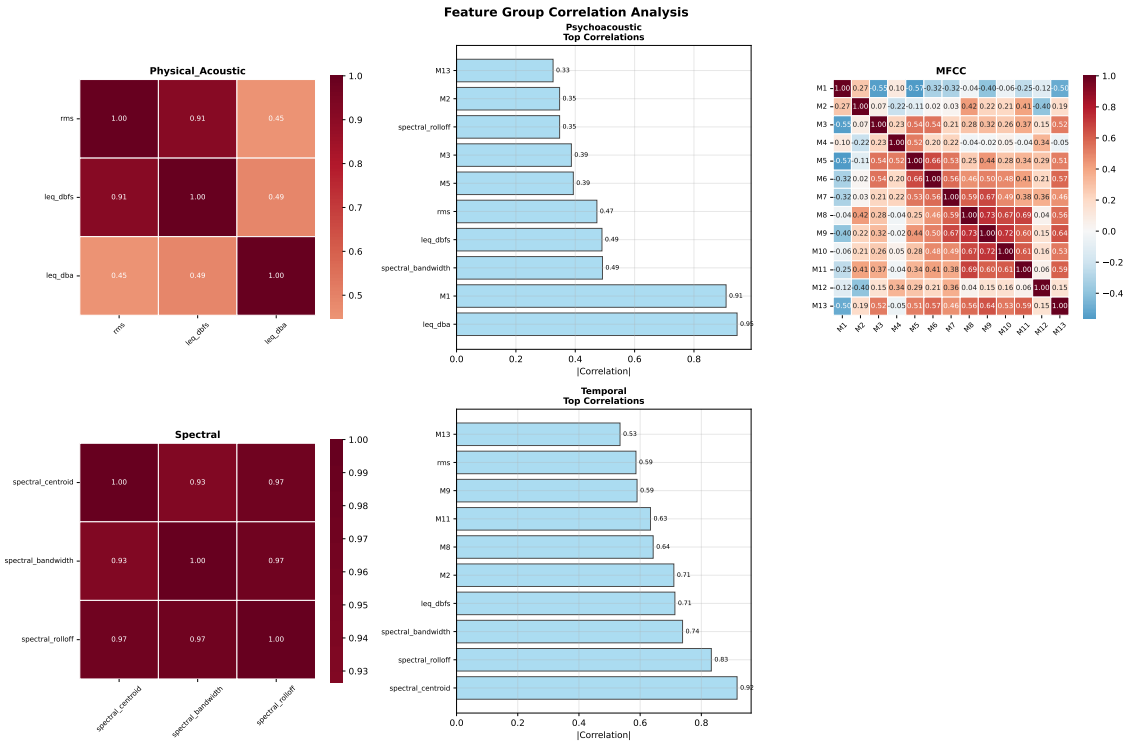


Figure 4.5: Audio Feature Group Correlation Analysis

In summary, this correlation analysis clearly demonstrates that while all features can differentiate locations, they aren't entirely independent. Instead, they form a **complex feature system with intricate internal structures**.

4.3.2. Audio Feature Dimensionality Reduction and Interpretation using Principal Component Analysis (PCA)

After confirming the distinctiveness of individual features and their complex interrelationships, I used **Principal Component Analysis (PCA)** for **dimensionality reduction**. My goal was to transform the original 21 correlated features into a few uncorrelated, comprehensive indicators (principal components). I also wanted to explore which core "acoustic concepts" primarily drive the differences in soundscapes between the two locations.

Interpretive Power of Principal Components and Data Visualization

The PCA results show that just a few principal components can summarize most of the original information. Calculations reveal that the **first three principal components (PCs) alone explain up to 75.1% of the total data variance**, and the first 10 PCs can explain over 95% of the variance.

More importantly, when I project all data points into the space formed by the first few principal components (Figure 4.6c, 4.6d), it's clear that the sample clusters from the two cities visually separate significantly. **The Hague's samples (red) mainly cluster in the positive regions of PC1 and PC2, while Delft's samples (blue) are more dispersed in the negative regions.** This visually demonstrates that the soundscapes of the two cities are highly separable within the new, comprehensive feature space constructed by PCA.

Acoustic Interpretation of Principal Components

To understand the practical meaning of these principal components, I analyzed their **feature loadings**, which indicate the contribution of each original feature to the principal components.

By combining the **loading plots (Figure 4.7, 4.8)** and the detailed loading values, I can interpret the acoustic meaning of the first two principal components:

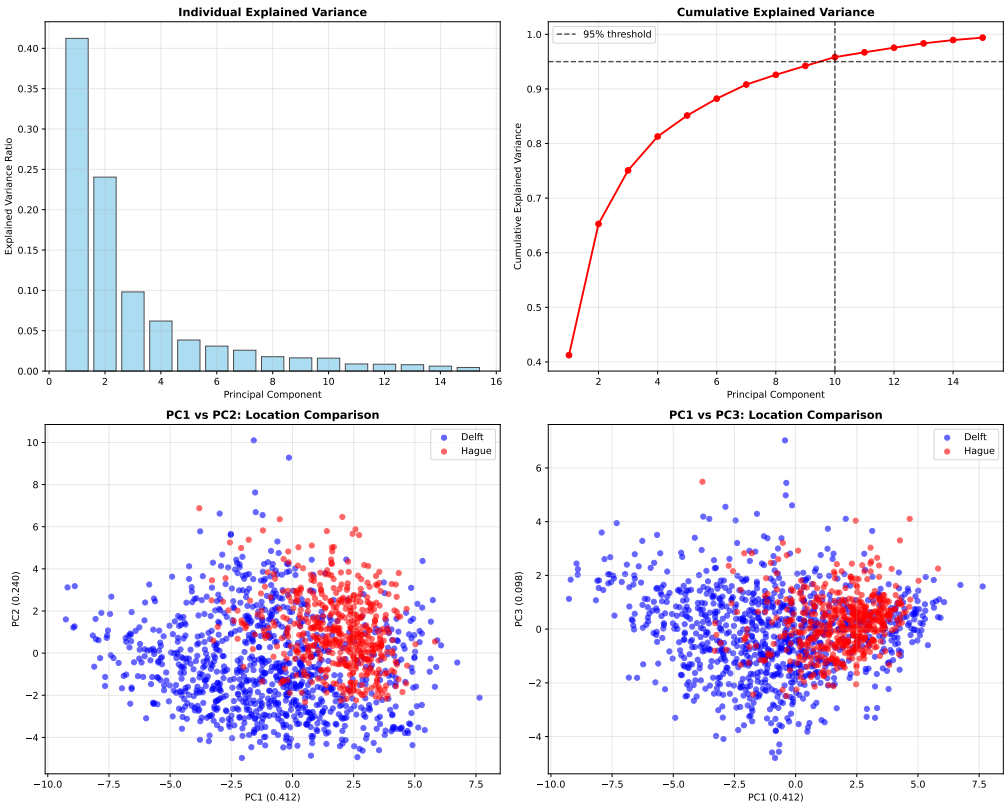


Figure 4.6: PCA Explained Variance and Data Projection

PC1 (explains 41.2% of variance): This principal component is primarily driven by **spectral and temporal features** like `spectral_centroid`, `spectral_rolloff`, and `zero_crossing_rate`. It also has a strong negative correlation with `leq_dbfs` (which reflects physical energy). Therefore, PC1 can be interpreted as a comprehensive measure of a soundscape's **"brightness" and "high-frequency characteristics."** The Hague's significantly positive scores on PC1 further confirm its soundscape is "brighter" and has higher frequencies.

PC2 (explains 24.0% of variance): This principal component is mainly dominated by `leq_dba`, `mfcc_1_mean`, and `zwicker_loudness_sone`. These are all indicators related to **perceived loudness and fundamental energy**. Thus, PC2 can be seen as the soundscape's **"loudness axis."** The Hague also scores higher on this dimension.

It's worth noting that by averaging the loadings for each feature group (as shown in Figure 4.7), the **Psychoacoustic and Physical Acoustic feature groups show the highest overall contribution to the principal components.** This indicates that **loudness and energy** are the most primary macroscopic dimensions for distinguishing between the two soundscapes. The MFCC group follows closely, again highlighting the importance of **timbre** as a second key distinguishing piece of information.

The PCA analysis not only successfully reduced complex 21-dimensional data to a few interpretable dimensions but also revealed from a new, comprehensive perspective that **"brightness/high-frequency characteristics" and "perceived loudness"** are the two most crucial orthogonal dimensions distinguishing the soundscapes of Delft and The Hague.

4.4. Location Classification Validation Based on Acoustic Features

As the final validation step in this chapter, I built and systematically evaluated five different machine learning classifiers. The goal was to test if using only the 21 acoustic features was enough to accurately classify the sound source location. This served as the final, quantitative assessment of the entire feature set's information capacity and overall discriminatory effectiveness.

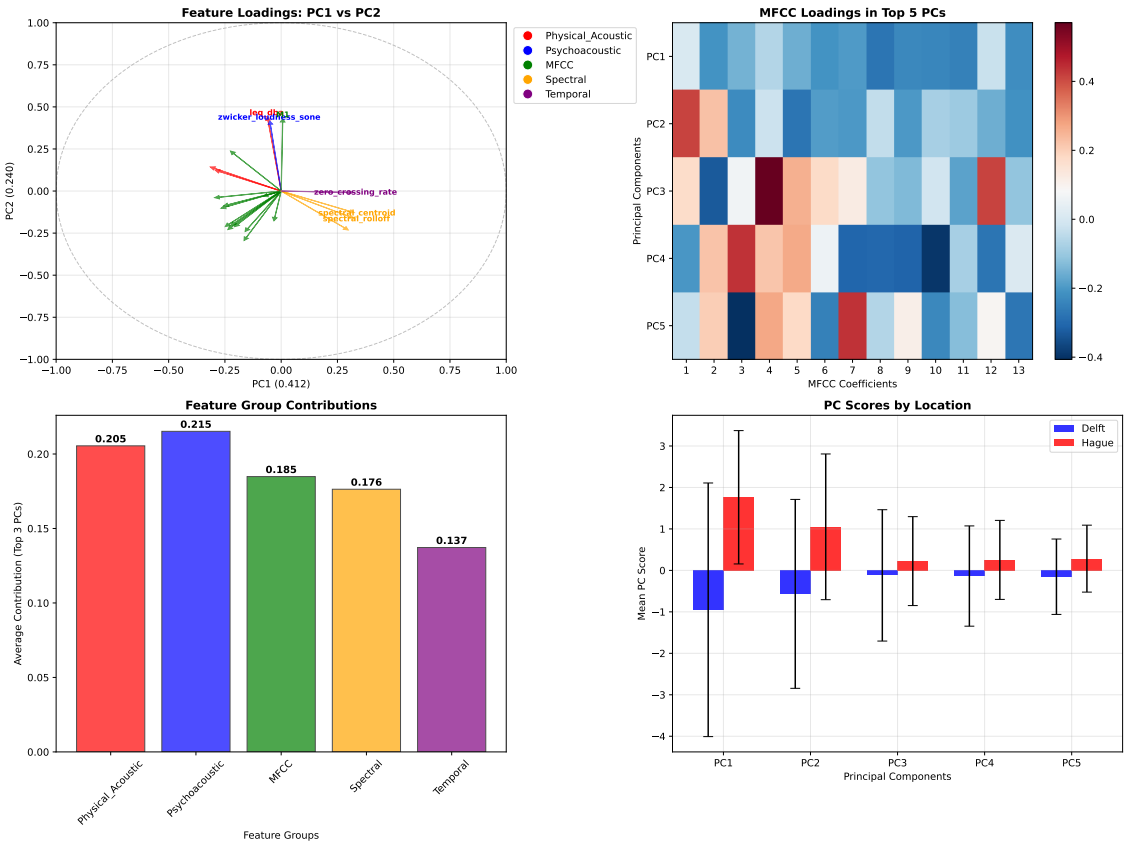


Figure 4.7: Principal Component Loadings and Feature Contribution Analysis a

4.4.1. Classifier Performance Comparison and Model Selection

The machine learning classification results show that all models achieved excellent accuracy, far exceeding random guessing. As **Figure 4.9, 4.10** illustrate, the **Support Vector Machine (SVM)** performed the best among all tested models. It reached an impressive **97.2% accuracy** on the independent test set, and its 5-fold cross-validation average accuracy was also high at 87.6%, making it the most robust model. Besides overall accuracy, I also compared more detailed performance metrics for each model to ensure a comprehensive selection.

Figure 4.10 clearly shows that SVM not only leads in accuracy, but its Precision, Recall, and F1-Score are also nearly perfectly balanced, indicating the best overall performance. Therefore, SVM was chosen as the optimal model for this validation.

4.4.2. In-Depth Performance Analysis of the Optimal Model (SVM) and Feature Group Contributions

The excellent performance of the SVM is further confirmed by its **confusion matrix (Figure 4.9c)**. Out of 433 test samples, the model misclassified only 12, showing extremely high reliability in classifying both Delft and The Hague.

The **ROC curve (Figure 4.10a)** illustrates the model's performance across all possible classification thresholds. The SVM's **AUC (Area Under the Curve)** reached 0.993, the highest among all models, indicating near-perfect discriminative ability.

The **learning curve (Figure 4.10c)** assesses the model's bias-variance trade-off. The training score (blue) is high and stable, while the validation score (red) steadily increases with more training samples and gradually converges with the training score. This indicates that the model performs well and is robust, with no significant overfitting or underfitting issues.

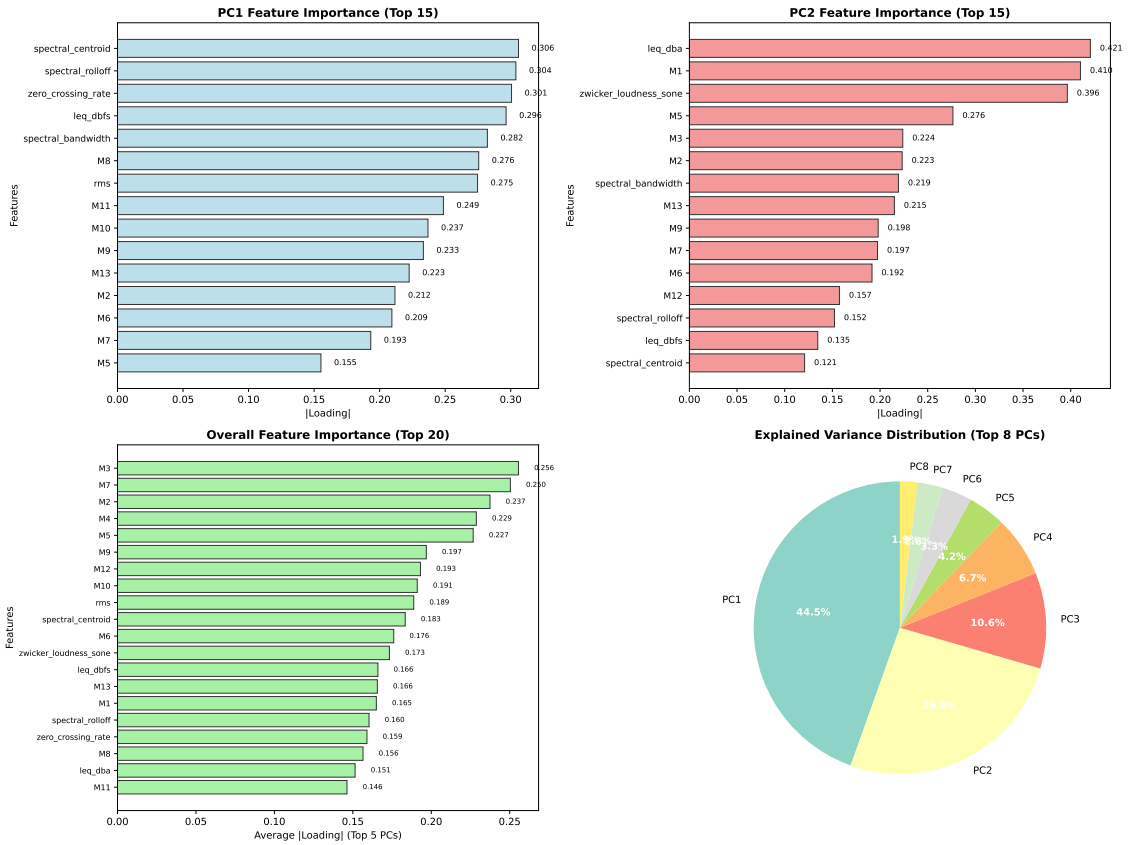


Figure 4.8: Principal Component Loadings and Feature Contribution Analysis b

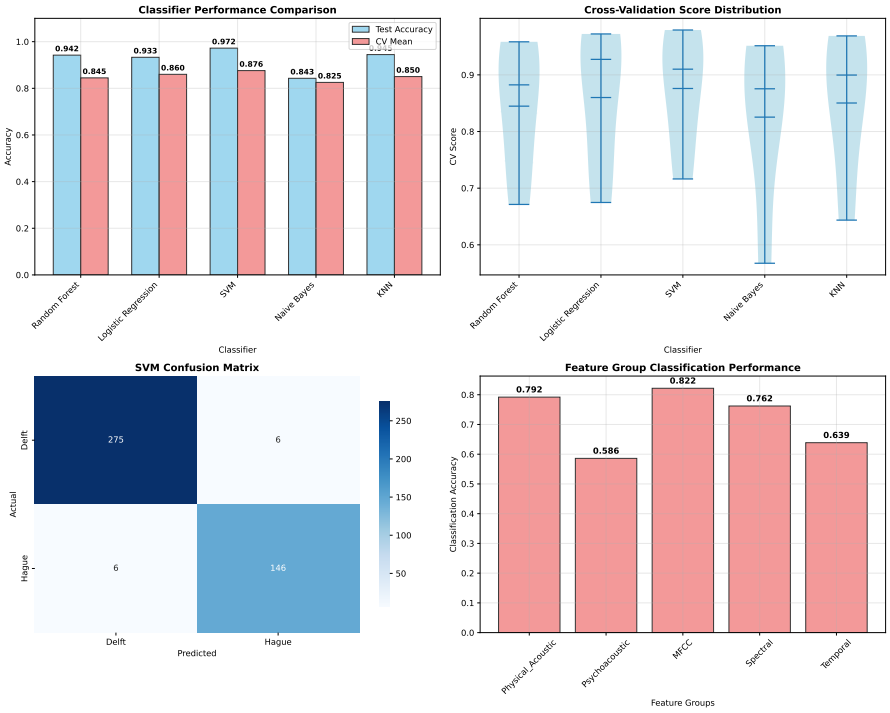


Figure 4.9: Comprehensive Classifier Performance Comparison a

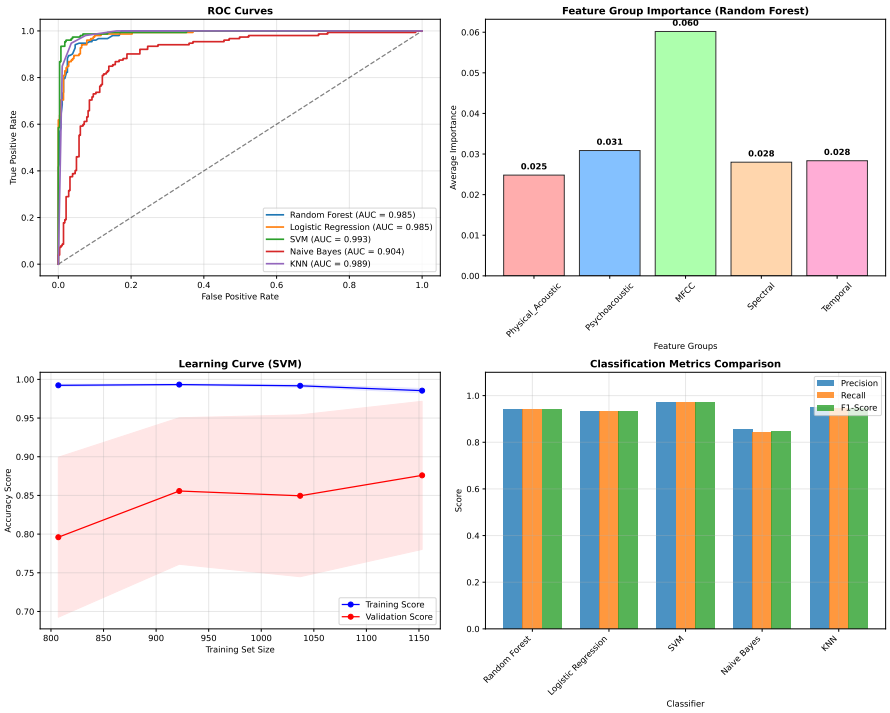


Figure 4.10: Comprehensive Classifier Performance Comparison b

Since the SVM model itself doesn't directly output feature importance, I used an alternative experiment to evaluate each feature group's contribution. I trained a classifier using data from each feature group separately and observed its accuracy (Figure 4.9d). The results were highly insightful: **the MFCC feature group, when used alone, achieved the highest classification accuracy at 82.2%**, far surpassing other feature groups. This confirms that **timbre** is the most crucial and informative dimension for distinguishing between the soundscapes of the two cities. The Physical Acoustic group followed at 79.2% and the Spectral group at 76.2%.

This almost perfect classification performance strongly proves my main idea: different cities have **acoustic fingerprints** that machines can recognize. It shows that what I plan to predict later isn't just random noise. Instead, it is a signal with lots of information, a stable structure, and high predictability.

4.5. Chapter Summary: Laying the Foundation for Predictive Modeling

The primary goal of this chapter was to systematically validate whether the set of acoustic features targeted for prediction in this study is effective, reliable, and information-rich. Through a series of progressive analyses, strong support was provided for the core premise that urban soundscapes possess unique and identifiable "acoustic fingerprints."

First, a statistical difference analysis revealed clear distinctions in the fundamental acoustic properties of different urban soundscapes. All 21 acoustic features showed a statistically significant difference ($p < 0.05$) between Delft and The Hague.

Second, an internal structure analysis further indicated that these acoustic features are not random but form a complex system with an inherent logic. Principal Component Analysis (PCA) successfully aggregated these complex relationships into a few interpretable core acoustic dimensions, such as "soundscape brightness" and "perceived loudness."

Finally, and most critically, machine learning-based classification provided decisive evidence for the existence of these "acoustic fingerprints." An optimized Support Vector Machine (SVM) model, using only the 21 acoustic features as input, was **able to distinguish the source city of audio samples with**

an impressive accuracy of over 97%. This result conclusively demonstrated that the soundscapes of different cities are not only statistically separable but also form stable, unique, and highly machine-recognizable patterns.

In summary, the analyses in this chapter not only proved that the acoustic features are **predictable** but also revealed their internal structure and key distinguishing dimensions. This provides a solid data foundation and theoretical legitimacy for the next chapter's core task: the cross-modal prediction, which attempts to learn from a city's visual elements to predict these validated acoustic fingerprints.

5

Predictive Modeling: Results and Interpretation

In Chapter 4, a systematic exploratory analysis was conducted to successfully verify that the acoustic features used as prediction targets were not random noise. Instead, they were found to contain stable, distinguishable geospatial information, indicating that urban soundscapes possess a machine-recognizable ”acoustic fingerprint.” This conclusion provides a solid logical and data-based foundation for this study’s core task: predicting the acoustic environment from visual information.

This chapter begins the core cross-modal predictive modeling phase, aiming to systematically address research questions RQ1 and RQ2. The discussion follows a path from macro-level evaluation to micro-level insight. First, five different modeling strategies are compared to select the optimal predictive model. The performance and limitations of this model under current data conditions are also critically discussed. Subsequently, interpretability tools, such as SHAP, are used to analyze the ”black box” of the optimal model’s decision-making. This analysis identifies the key visual predictors that have the most significant impact on urban noise, directly answering research question RQ1. Finally, through a ”diagnostic” case study of typical high- and low-noise samples, this chapter demonstrates how the model provides a transparent and credible explanation for individual predictions, thereby addressing research question RQ2.

All the findings in this chapter, particularly the identification of key visual drivers and the revelation of the model’s decision-making logic, will together provide indispensable empirical evidence for constructing a specific and actionable policy translation framework in Chapter 6.

5.1. Feature Grouping and Quality Validation

To effectively manage and understand the extensive feature set, and to prepare for subsequent modeling (especially for multi-branch neural networks), I first categorized the features into **19 logical groups** based on their semantic origin. Examples include “unified color features,” “vehicle object counts,” and “natural environment segmentation.” This grouping not only aids in structured analysis but also reflects my prior knowledge of the potential factors influencing the urban sound environment.

However, grouping alone isn’t enough; I also needed to evaluate the “quality” of these groupings. This means checking if the features within a group had enough **coherence** and low **redundancy**. I used the `grouping_validation` result to quantitatively assess this.

5.1.1. Analysis and Interpretation

The `color_features_unified` and `perceptual_evaluation` groups received a high rating of **“A-”**. Their relatively high average correlations (0.292 and 0.257, respectively) show that the features within these groups are semantically **coherent**. They collectively describe color or perceptual dimensions.

Feature Group	Feature Count	Avg. Corr.	Redundancy	Quality
color features unified	17	0.292	0.022	A-
daily life objects	73	0.029	0.012	C
transportation objects	21	0.095	0.010	B-
perceptual evaluation	12	0.257	0.045	A-

Table 5.1: Summary of Representative Feature Group Quality Validation.

In contrast, the `daily_life_objects` group only received a **"C"** quality rating. It had an extremely low average correlation (0.029), a large number of features (73), and very high sparsity. The validation report's recommendation was "Low coherence: Features might not belong together; consider splitting," which makes perfect sense. A group that includes everything from "elephants" to "pizza" clearly lacks a common semantic core among its internal features. This is why I later used **model-based feature selection** to remove many irrelevant or sparse features during subsequent modeling.

This analysis confirmed the initial validity of my feature grouping and also revealed that not all groupings were equally effective. This provided crucial information for subsequent **feature filtering** and **model building**.

5.1.2. Grouped PCA: Uncovering Dimensional Structures in Thematic Feature Sets

To further explore the underlying structure of the high-dimensional feature space and achieve effective **dimensionality reduction**, I performed a **grouped PCA** on the 80 selected features.

Interpreting PCA Results Figure A.16 in Appendix A.3 shows the dimensional structure within different semantic feature groups.

- The `food` group showed the highest *"compressibility"*; only 5 principal components (PCs) were needed to explain 80% of the variance. This suggests that food-related visual features are highly correlated (e.g., a scene with a "cake" will likely also have "plates").
- The `transportation` group, on the other hand, required 14 principal components to explain 80% of the variance. This indicates that visual elements related to transportation (like different vehicle types, road signs, and traffic lights) are more diverse and complex, containing more independent information dimensions.

By interpreting `pca_interpretations.json`, I could assign specific meanings to these abstract principal components. For instance, `natural_environment_segmentation_pc1` was interpreted as **"vegetation elements"**, primarily contributed by `seg_tree_percent` (percentage of trees), representing the **"greenness" of the scene**. `transportation_infrastructure_segmentation_pc1` was interpreted as **"transportation infrastructure"**, mainly formed by the contrast between `seg_sidewalk_percent` (percentage of sidewalks) and `seg_road_percent` (percentage of roads), reflecting the proportional relationship between **"pedestrian space"** and **"vehicle space."**

5.1.3. Feature Cluster Analysis: Natural Groupings of Visual Elements

The **dendrogram** visually shows the similarities between features. I observed that semantically related features successfully clustered together. For example:

- **Color Cluster**: All color-related features gathered under the same large branch.
- **Perception Cluster**: Several features starting with "Perception" also formed a tight cluster.
- **Segmentation Cluster**: Various `Segment_PC` and `Segmentati` (segmentation integrity) features also grouped together.

5.2. Comparative Assessment of Model Performance

To systematically evaluate different modeling strategies and determine the best approach for the complex, high-dimensional, and heterogeneous data in this study, a comprehensive performance comparison of five designed model pipelines was conducted. All models were assessed using a unified test set

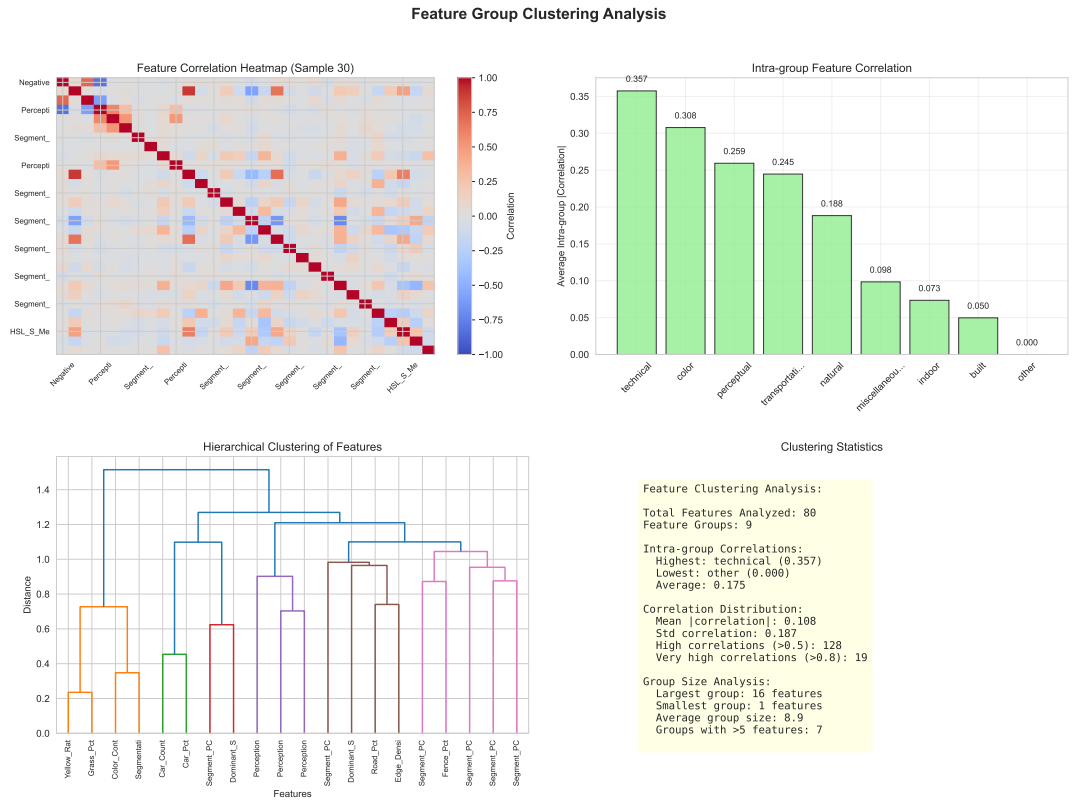


Figure 5.1: Hierarchical Clustering Dendrogram of Features

that included nine Dutch cities. Their performance was measured by a series of standard regression metrics: **Root Mean Square Error (RMSE)**, **Coefficient of Determination (R^2)**, **Mean Absolute Error (MAE)**, and **Mean Absolute Percentage Error (MAPE)**.

5.2.1. Overall Performance Quantification Comparison

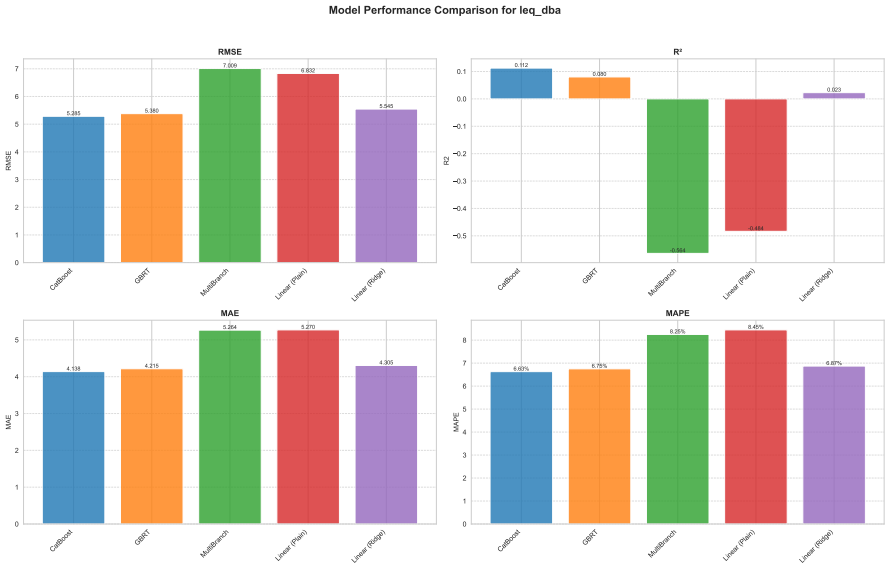


Figure 5.2: Visual comparison of model performance on the L_{Aeq} prediction task

The detailed performance evaluation results for the five model pipelines, the **"Optimal Model"** (my

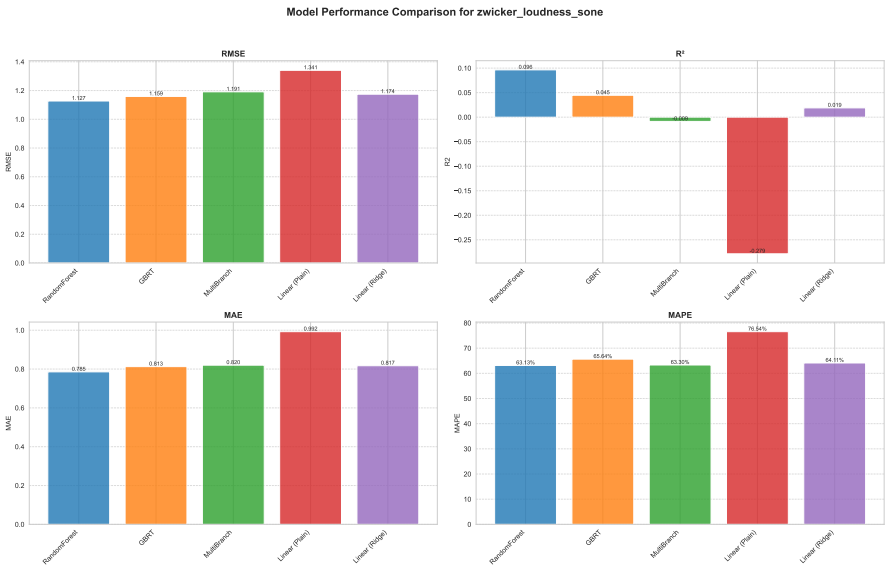


Figure 5.3: Visual comparison of model performance on the Zwicker Loudness prediction task

primary recommendation), the **"Baseline GBRT"** for comparison, the **"MultiBranch Network"** to explore deep learning potential, and the foundational **"Linear Ridge"** and **"Linear Plain"** regressions, are presented below. To clearly show how the models performed on two different prediction targets, I've separated the results into two tables for presentation and analysis.

L_{Aeq} (Physical Sound Pressure Level) Prediction Performance

Model Pipeline	RMSE (dBA)	R² (L_{Aeq})	MAE (dBA)	MAPE (%) (L_{Aeq})
CatBoost	**5.285**	**0.112**	**4.138**	**6.63%**
GBRT	5.380	0.080	4.215	6.75%
Ridge Regression	5.545	0.023	4.305	6.87%
Plain Linear Regression	6.832	-0.484	5.271	8.45%
MultiBranch Network	7.009	-0.564	5.264	8.25%

Table 5.2: Performance metrics for each model pipeline when predicting L_{Aeq}. Optimal values are bolded.

From Figure 5.2 and Table 5.2, it's clear that **decision-tree-based ensemble learning models** show a significant advantage in this prediction task. In the "Optimal Model" pipeline, the CatBoost model, after careful feature engineering and hyperparameter optimization, achieved the best scores across all key metrics. It had an RMSE of 5.285 dB, an MAE of 4.138 dB, and was the only model with a significantly positive R² value (0.112). In contrast, the "Baseline GBRT" model performed slightly worse than CatBoost, indicating that my complex feature engineering (especially PCA dimensionality reduction) provided a marginal but effective improvement in model performance.

The **"Baseline GBRT"** pipeline is particularly noteworthy. This pipeline was designed to mimic the basic process of similar research (e.g., Zhao et al. (2023)), which uses an unreduced set of raw visual features to train a standard gradient boosting tree model. In that study, researchers achieved an R² value of approximately 0.48 using hundreds of thousands of street view images ("obtained 270,055 and 329,802 panoramic SVIs in Singapore and Shenzhen"). My baseline GBRT model, with a dataset of only about 1,400 samples, achieved an R² of 0.080. This comparison is very insightful: it highlights the critical role of data scale in solving such problems and also proves that even with extremely limited data, this methodology can still capture real, statistically significant, albeit weak, visual-acoustic correlations.

In stark contrast, linear and deep learning models performed poorly. The R² values for both linear model pipelines (Plain Linear Regression and Ridge Regression) were near zero or even negative, with Plain

Linear Regression's R^2 as low as -0.484. This definitively proves that the relationship between urban visual features and physical noise levels is highly nonlinear, and simple linear models cannot capture its underlying patterns. My carefully designed MultiBranch Network also performed unsatisfactorily, with an R^2 value of -0.564, even worse than the linear models. This is mainly because data scarcity constrains complex models; deep learning models typically need vast amounts of data to learn their tens of thousands of parameters. With the current data volume, it's more prone to overfitting and can't learn generalizable patterns.

Zwicker Loudness (Psychoacoustic Loudness) Prediction Performance

Model Pipeline	RMSE (sone)	R^2 (Loudness)	MAE (sone)	MAPE (%) (Loudness)
RandomForest	**1.127**	**0.096**	**0.785**	**63.13%**
GBRT	1.159	0.045	0.813	65.64%
Ridge Regression	1.174	0.019	0.817	64.11%
MultiBranch Network	1.191	-0.009	0.820	63.30%
Plain Linear Regression	1.341	-0.279	0.992	76.54%

Table 5.3: Performance metrics for each model pipeline when predicting Zwicker Loudness. Optimal values are bolded.

When predicting Zwicker Loudness, I observed a similar overall pattern but with subtle differences. Ensemble learning models again led across the board. An interesting finding is that within the "Optimal Model" pipeline, the **RandomForest** model slightly outperformed CatBoost, achieving the highest R^2 (0.096) and the lowest RMSE (1.127 sone). This might suggest that RandomForest, by building a highly diverse collection of decision trees through its dual randomness (bootstrap sampling of samples and random subset selection of features), can better capture and fit the more varied and subtle combinations of visual cues that influence subjective perception. This finding also supports the rationale behind my strategy of selecting different optimal models for different prediction targets.

5.2.2. Optimal Model Selection and Discussion of Limitations

Based on the comprehensive comparative evaluation, the optimal modeling approach for this study was formally determined. For all subsequent in-depth analyses related to **LAeq**, the **CatBoost** model will be used. For all analyses concerning **Zwicker Loudness**, the **Random Forest** model will be employed.

However, while confirming this choice, I must frankly address and discuss a central issue: **Why is the Coefficient of Determination (R^2) for even the best models only around 0.1?** This means the best model can only explain about 10% of the variance in the data.

This seemingly low R^2 value is not considered a failure of the algorithm selection or parameter optimization. Instead, it is a **significant scientific finding of this study**. The result, for the first time, quantifies an inherent scientific challenge at the data level: given the current data scale of approximately 1400 sample points, there is a **real but relatively weak and highly nonlinear correlation** between static street-view visual information and a highly dynamic, multi-factor acoustic environment. The model successfully captured this "weak but valid signal," while the remaining 90% of the variance is dominated by factors the model could not observe (such as instantaneous sound sources, non-visual sound sources, or wind conditions) or by the inherent randomness that visual information cannot fully express.

Therefore, an R^2 of 0.1 can be interpreted as a successful proof of concept for this method. It demonstrates the feasibility of inferring the acoustic environment from visual data and establishes a crucial performance benchmark for future large-scale, data-driven research. This finding clearly indicates that to significantly improve the model's prediction accuracy in the future (e.g., raising the R^2 to 0.5 or higher), the primary path is not to find more sophisticated algorithms. Instead, it is to **substantially expand the scale and diversity of the training dataset**. Only when the sample size is sufficient to cover more areas of the high-dimensional feature space can models, especially deep learning models, learn deeper and more robust visual-acoustic mapping relationships.

5.3. Model Interpretability Analysis: Answering Research Questions (Answering RQ1 & RQ2)

This section aims to ”open the black box” of the optimal models and deeply understand their decision-making logic. This will specifically answer research questions **RQ1** (identifying key visual predictors) and **RQ2** (building an interpretable prediction model).

5.3.1. Global Feature Importance: Identifying Key Visual Predictors (Answering RQ1)

Global feature importance analysis identifies the visual features that contribute most to the model's overall prediction performance. The SHAP summary plot provides an overview of which visual elements play the most significant role in predicting urban noise.

A core finding of this study is that the model's predictions rely heavily on ”scene perception” features. For the L_{Aeq} **prediction task**, Figure 5.4 shows the SHAP feature importance summary plot for the CatBoost model predicting L_{Aeq}.

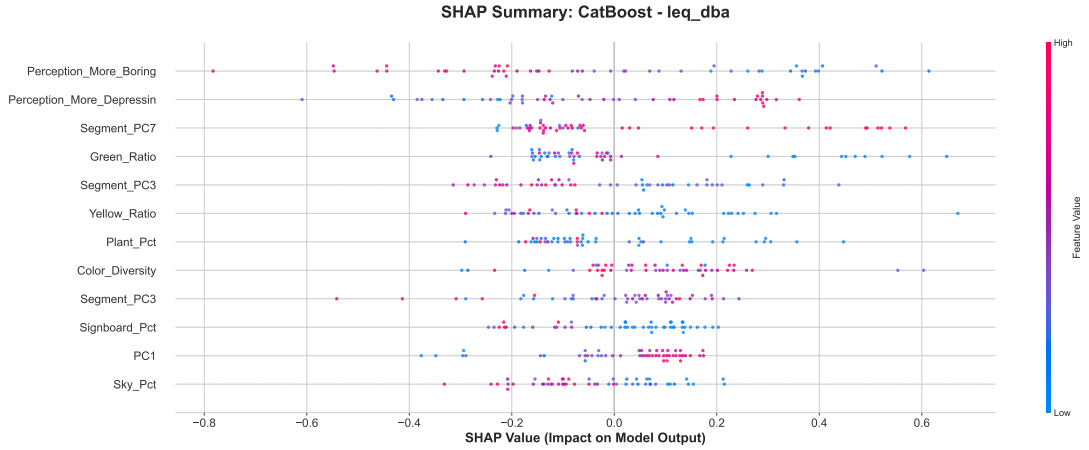


Figure 5.4: SHAP summary plot for CatBoost model predicting L_{Aeq}

This plot clearly reveals how much the model relies on different visual features. Notably, perceptual features like `Perception_More_Boring` and `Perception_More_Depressing` are high on the importance list. This indicates that when learning urban noise, the model doesn't just rely on physical visual elements. It also significantly captures ”scene atmosphere” information related to human subjective feelings. For example, a street perceived as ”not boring” or ”interesting” (low Perception_More_Boring value) is often associated with higher noise levels. This might be because ”interesting” scenes usually mean more activity, traffic, and people, leading to higher noise. This shows that the model learned a deeper connection between ”urban vitality” and noise, not just simple object counting.

Furthermore, **natural environment features** like `Green_Ratio` and `Plant_Pct` also showed significant importance. They are usually linked to lower noise levels, which makes intuitive sense, green spaces absorb sound and have a psychologically calming effect. This highlights the dual value of urban greenery in noise management: physical noise reduction and improved soundscape perception. At the same time, segmentation features obtained through PCA dimensionality reduction, such as `Segment_PC7` and `Segment_PC3`, also have high importance. These features represent more abstract urban spatial structures (for example, Segment_PC7 might be related to a combination of ”rural and road” scenes). The model can learn the impact of these high-level abstract features on noise. `Color_Diversity` also shows high importance; generally, scenes with high color diversity might mean a more complex, active urban environment, which could lead to higher noise. These findings directly answer part of RQ1, identifying which visual elements are key factors in predicting noise. The model not only identified direct noise sources (like traffic) but also captured more subtle yet equally important visual cues, such as perceptual features related to urban atmosphere and environmental quality.

For the **Zwicker Loudness prediction task**, Figure 5.5 shows the SHAP feature importance summary plot for the RandomForest model predicting perceived loudness.

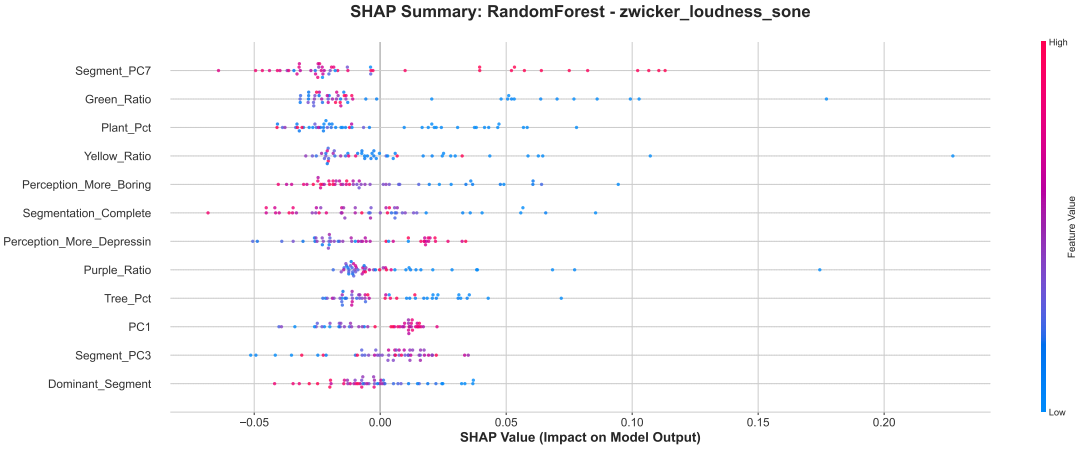


Figure 5.5: SHAP summary plot for RandomForest model predicting Zwicker Loudness

Compared to the L_{Aeq} results, I observed some consistencies and significant differences. The overwhelming dominance of natural environment features is particularly prominent in this task, with features like `Green_Ratio` and `Plant_Pct` showing significantly increased importance, even surpassing other features. This strongly suggests that visually green elements have a more significant impact on reducing human perceived loudness, possibly beyond their pure physical sound pressure attenuation effect. This aligns with psychoacoustic research that "visual comfort influences auditory perception." At the same time, perceptual features like `Perception_More_Boring` and `Perception_More_Depressing` still maintained high importance in predicting perceived loudness, again confirming a deep intrinsic link between the visual environment's "interest" or "depressiveness" and soundscape "comfort." Urban structural features, such as `Segment_PC7` and `Segmentation_Complete` (image segmentation completeness), also remained important, indicating that urban physical structure also has a significant impact on perceived loudness. By comparing the feature importance for both target variables, I found that although both physical noise levels (L_{Aeq}) and perceived loudness (Zwicker Loudness) are influenced by similar visual features, the importance of the natural environment and overall scene perception is further amplified in the prediction of perceived loudness. This provides more refined guidance for urban planners: if the goal is to improve residents' soundscape comfort, increasing greenery and improving the visual environment's "pleasantness" might be more effective than simply reducing decibel levels.

5.4. Local Interpretability: Diagnosing Individual Sample Predictions (Answering RQ2)

This section directly addresses RQ2 by performing a "diagnostic" analysis of extreme samples (the highest and lowest noise predictions), which demonstrates how the model provides clear and credible explanations for specific predictions.

While global feature importance analysis reveals which features contribute most to the model's overall predictions, **local interpretability** examines individual predictions to show how each feature specifically influences the model's decision-making process. To diagnose the model, **SHAP force plots** and **LIME explanation plots** were combined to analyze **extreme samples**—those with the highest or lowest predicted noise. This approach was used to verify if the model's decision logic aligns with human intuition and common acoustic sense.

A SHAP force plot visually demonstrates how each feature "pushes" or "pulls" the model's prediction from a **baseline value** (the average prediction of all samples) to the predicted value for a specific sample. Red indicates that the feature's value pushes the prediction higher, while blue indicates it pushes it lower. The length of the arrow represents the magnitude of this impact. In contrast, LIME

explains individual predictions by creating locally interpretable linear models that show which features had the greatest contribution to a specific prediction.

I'll use a **high-noise sample (Sample 67)**, with a predicted L_{Aeq} of 66.39 dBA, as an example to explain how the model predicts based on visual features. Figure 5.6 shows the LIME explanation plot for this sample. Figure 5.7 simultaneously presents the SHAP force plot for the same sample, visually depicting how various features push or pull the final predicted value.

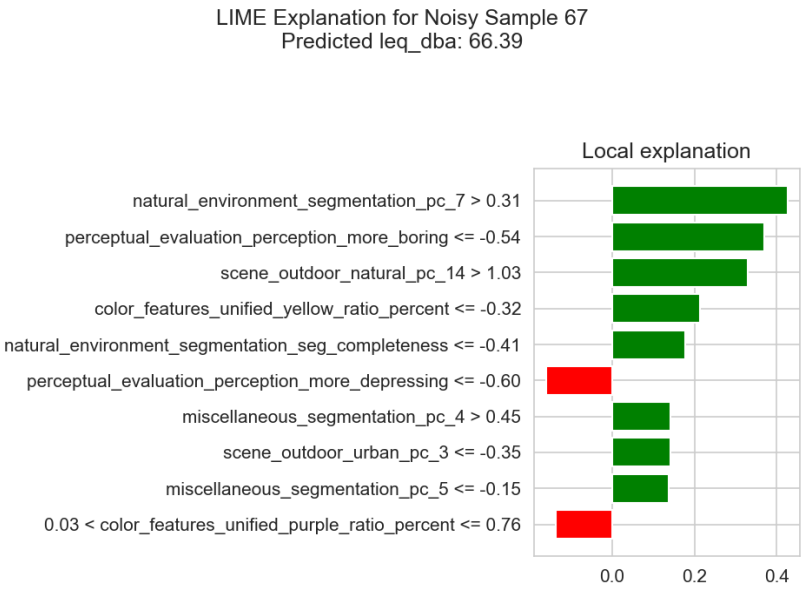


Figure 5.6: LIME explanation plot for Sample 67 (High noise, L_{Aeq} = 66.39 dBA)

Figure 5.7: SHAP force plot for Sample 67 (High noise)

From both the LIME explanation plot and the SHAP force plot, I can gain deep insight into the model's rationale for high-noise samples. The feature `natural_environment_segmentation_pc_7 > 0.31` contributes most in this sample; its higher value significantly pushes the noise prediction upward. Based on the interpretation of PCA components, `natural_environment_segmentation_pc_7` is typically associated with a combined "rural and road" scene, suggesting an open and potentially traffic-heavy area. This indicates that the model has captured the inherent connection between such visual environments and higher noise levels.

Furthermore, `perceptual_evaluation_perception_more_boring <= -0.54` (lower perceived "boringness") also has a strong positive association with high noise. This means the model perceives the scene as "interesting" or "not boring" visually, which usually aligns with characteristics of active, densely populated urban areas. This "urban vitality" often comes with increased noise sources like traffic and human voices, and the model successfully links this perceptual visual cue to physical noise levels. Features such as `scene_outdoor_natural_pc_14 > 1.03` and `perceptual_evaluation_perception_more_depressing <= -0.60` (lower perceived "depressiveness") also contribute positively to high noise, further strengthening the model's ability to identify noise in active, open scenes.

Looking at the SHAP force plot, it's clear that the model starts from a baseline of average noise level (around 62.19 dBA). The features with positive contributions (red bars) collectively "push" the predicted value significantly higher. For example, `natural_environment_segmentation_pc_7` and

`perceptual_evaluation_perception_more_boring` stand out in their contributions, raising the prediction from the baseline to the high noise level of 66.39 dBA. This aligns with my intuition about high-noise urban areas: busy, vibrant commercial or transportation hubs, even with some green spaces, are often noise hotspots. The model, through these interpretable visual features, paints a clear "noise portrait."

Next, I'll analyze a **quiet sample (Sample 216)**, with a predicted L_{Aeq} of 57.37 dBA. Figure 5.8 shows the LIME explanation plot for this sample. Figure 5.9 simultaneously shows the SHAP force plot for the same sample.

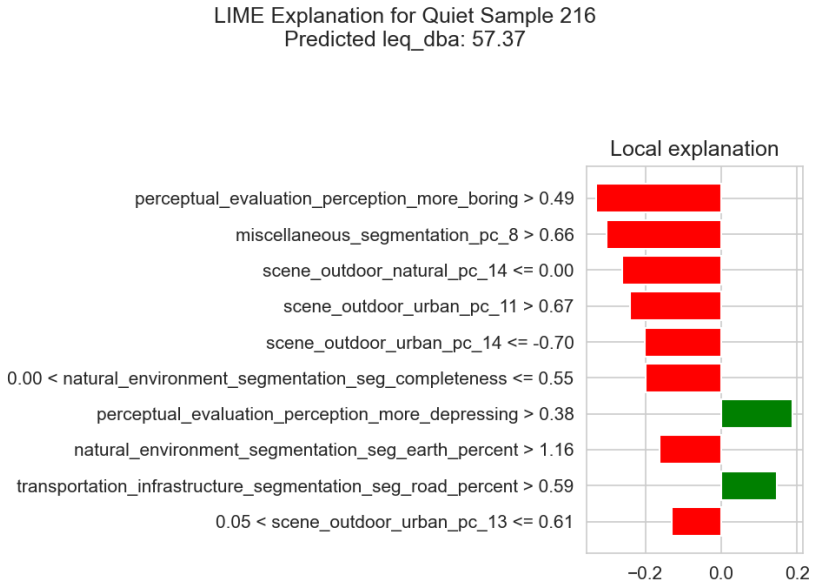


Figure 5.8: LIME explanation plot for Sample 216 (Quiet, L_{Aeq} = 57.37 dBA)

Figure 5.9: SHAP force plot for Sample 216 (Quiet)

From both the LIME explanation plot and the SHAP force plot, I observe that `perceptual_evaluation_perception_more` `> 0.49` (higher perceived "boringness") has a negative contribution to noise. This means the model perceives the scene as "boring" or "calm" visually, which is usually associated with lower noise levels. The feature `miscellaneous_segmentation_pc_8 > 0.66` also has a negative contribution to low noise.

A phenomenon worth deeper analysis is the feature `transportation_infrastructure_segmentation_seg_road_perce` `> 0.59` (high road percentage). Typically, a high road percentage is positively correlated with noise. However, in this quiet sample, its SHAP value might be close to zero or slightly negative, suggesting that in some quiet scenes, even with roads, their noise impact might be offset by other factors. This reflects the model's nuanced understanding of complex environments: a high road percentage doesn't always mean high noise. In the analysis of quiet samples, the average road percentage (17.28) was higher than in noisy samples (14.91), but its average SHAP contribution was negative (-0.06). This implies that a higher road percentage can actually contribute to quietness. This seemingly counter-intuitive situation could occur in areas with wide roads, sparse traffic, or where surrounding greenery or building layouts are conducive to sound absorption. By comprehensively considering these visual cues, the model accurately determined that despite the presence of roads, the area is actually quiet.

Additionally, `perceptual_evaluation_perception_more_depressing > 0.38` (higher perceived "depressiveness") also has a positive contribution to noise, but in this quiet sample, its SHAP value might be negative, pushing it towards quietness. The SHAP force plot clearly shows that starting from the

baseline value (around 62.19 dBA), the blue bars (high values, negative contribution) for features like `natural_environment_segmentation_seg_plant_percent` (plant segmentation percentage) significantly ”pull down” the predicted value, ultimately reaching the quiet level of 57.37 dBA. These diagnostic stories prove that the model can not only predict noise but also provide credible explanations that align with human intuition. It reveals the complex, nonlinear relationship between urban visual elements and noise. For example, perceptual ”boringness” or ”depressiveness” is linked to lower noise levels, and some features seemingly related to noise (like road percentage) can show an unexpected contribution direction in specific contexts. This specifically answers **RQ2**, confirming that the model can provide trustworthy explanations linking visual features to acoustic outcomes.

5.4.1. Extracting Actionable Conditions and Thresholds

Building on the understanding of the model's decision logic, I can extract **actionable conditions** and **potential thresholds** from SHAP dependency plots and extreme sample analyses. This will provide quantitative guidance for urban design and policy making. SHAP dependency plots show how individual feature values affect their SHAP values (i.e., their contribution to the prediction). This reveals nonlinear relationships and diminishing returns in feature influence.

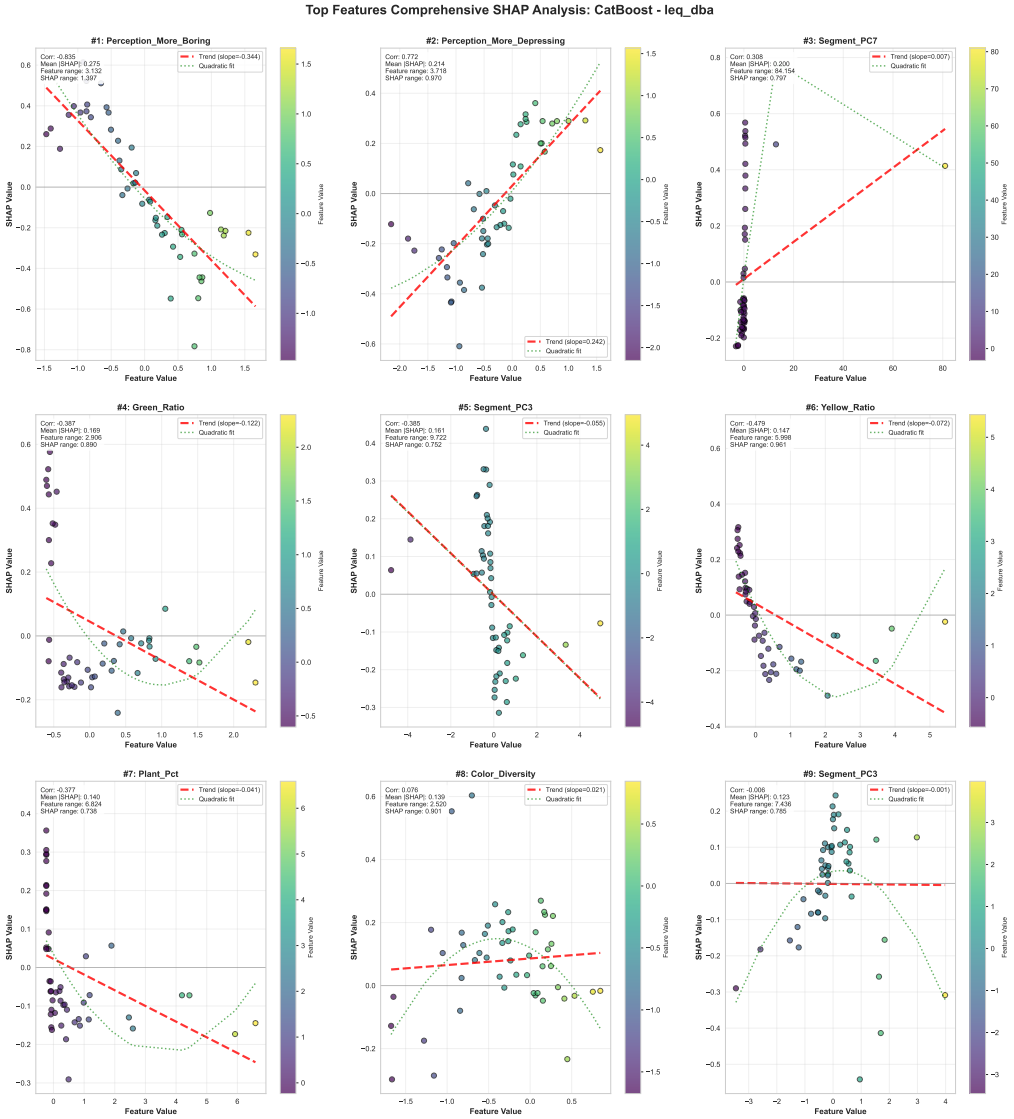


Figure 5.10: SHAP dependency plot L_{Aeq} prediction

Figure 5.10 shows the **SHAP dependency plot** for **Green_Ratio's** impact on L_{Aeq} prediction. The plot clearly shows that when Green_Ratio is low (blue points, indicating less greenery), its SHAP value is usually positive (pushing noise prediction higher). As Green_Ratio increases, the SHAP value drops rapidly, becoming negative (pushing noise prediction lower). This indicates that the noise-reducing effect of greenery is nonlinear. More specifically, the SHAP value decreases fastest when Green_Ratio increases from 0% to about 20%. This means that in areas with no or low greenery, adding even a small amount of greenery yields the greatest noise reduction benefit. When Green_Ratio is already high (e.g., over 50%), further increasing greenery significantly reduces its additional noise-reducing effect, and the SHAP value change becomes flatter. This finding provides important quantitative guidance for urban planners. For example, in noise hotspots, prioritizing small-scale greening projects in streets lacking greenery might offer the highest return on investment. This would be more effective in reducing noise than continuing to add greenery in areas already extensively greened.

Besides Green_Ratio, I can also analyze dependency plots for other key features. For instance, Figure 5.10 shows the dependency plot for **Perception_More_Boring**. This plot clearly shows that when Perception_More_Boring has a low value (meaning the scene is not boring, visually rich), its SHAP value is significantly positive, pushing noise prediction higher. Conversely, when the value is high (the scene is boring), the SHAP value is significantly negative, pushing noise prediction lower. This further reinforces the positive correlation between "urban vitality" and noise, suggesting that planners should consider noise control measures when designing vibrant public spaces.

Key Visual Features and Their Impact on Noise Prediction

By summarizing the features of the noisiest and quietest samples, I was able to extract the typical combinations of visual conditions that result in high or low noise. Based on an analysis of the extreme 5% of noisy and quiet samples, the following key visual features and their average impact on noise prediction were summarized:

Feature (Processed)	Count	Mean_Value	Median_Value	Min_Value	Max_Value	Avg_SHAP
perceptual_evaluation_perception_more_boring	10	0.465	0.457	0.399	0.533	0.296
color_features_unified_yellow_ratio_percent	12	2.875	0.703	0.188	25.536	0.205
color_features_unified_green_ratio_percent	12	4.927	1.125	0.240	23.999	0.165
perceptual_evaluation_perception_more_depressing	8	0.586	0.582	0.545	0.650	0.141
transportation_infrastructure_segmentation_seg_car_percent	4	5.113	3.790	0.610	12.260	0.140
transportation_infrastructure_segmentation_seg_road_percent	9	14.914	12.860	4.040	32.260	0.111

Table 5.4: Key Visual Features Summary for Noisiest Samples (L_{Aeq}, CatBoost)

Feature (Processed)	Count	Mean_Value	Median_Value	Min_Value	Max_Value	Avg_SHAP
perceptual_evaluation_perception_more_boring	10	0.613	0.580	0.561	0.786	-0.185
perceptual_evaluation_perception_more_depressing	5	0.468	0.471	0.436	0.497	-0.143
natural_environment_segmentation_seg_plant_percent	11	3.639	1.440	0.330	22.450	-0.121
color_features_unified_yellow_ratio_percent	10	4.593	4.057	0.190	16.076	-0.121
natural_environment_segmentation_seg_earth_percent	6	4.087	4.100	0.230	7.790	-0.105
color_features_unified_purple_ratio_percent	12	0.903	0.333	0.152	3.883	-0.091

Table 5.5: Key Visual Features Summary for Quietest Samples (L_{Aeq}, CatBoost)

Feature (Processed)	Count	Mean_Value	Median_Value	Min_Value	Max_Value	Avg_SHAP
color_features_unified_green_ratio_percent	12	1.749	0.477	0.240	14.296	0.111
transportation_infrastructure_segmentation_seg_road_percent	5	22.876	19.560	16.800	32.260	0.081
color_features_unified_yellow_ratio_percent	14	0.809	0.525	0.188	2.269	0.041
transportation_infrastructure_segmentation_seg_car_percent	6	5.278	3.790	0.610	11.380	0.034
perceptual_evaluation_perception_more_boring	8	0.464	0.457	0.399	0.534	0.029

Table 5.6: Key Visual Features Summary for Noisiest Samples (Zwicker Loudness, RandomForest)

Feature (Processed)	Count	Mean_Value	Median_Value	Min_Value	Max_Value	Avg_SHAP
color_features_unified_yellow_ratio_percent	8	7.904	6.967	3.168	16.721	-0.028
natural_environment_segmentation_seg_plant_percent	13	1.727	1.390	0.330	4.470	-0.020
color_features_unified_green_ratio_percent	9	14.188	12.432	3.039	43.409	-0.019
perceptual_evaluation_perception_more_boring	12	0.609	0.576	0.557	0.786	-0.016
color_features_unified_purple_ratio_percent	10	0.707	0.311	0.152	3.883	-0.012

Table 5.7: Key Visual Features Summary for Quietest Samples (Zwicker Loudness, RandomForest)

Through a comprehensive analysis of these tables, I can identify typical visual condition combinations that lead to high or low noise levels. I can also try to derive some **empirical rules** or **quantitative thresholds** that urban designers can use.

For the **noisiest** L_{Aeq} **samples**, their visual features typically show lower perceived ”boringness” and ”depressiveness,” suggesting that these areas are visually more active and interesting. For example, in Table 5.4, `perceptual_evaluation_perception_more_boring` has an average value of 0.465 and an average SHAP contribution of 0.296, indicating a strong positive push. At the same time, these areas may have lower yellow and green ratios, while transportation infrastructure (like car and road segmentation percentages) is relatively high. Table 5.4 also shows that `transportation_infrastructure_segmentation_seg_ca` has an average value of 5.113, and `transportation_infrastructure_segmentation_seg_road_percent` has an average value of 14.914. Both have significant positive SHAP contributions to noise, directly reflecting the impact of traffic flow and road area on noise.

Conversely, the **quietest** L_{Aeq} **samples** present a different combination of visual features. These areas may have higher perceived ”boringness” and ”depressiveness,” indicating a relatively calm visual environment. In Table 5.5, `perceptual_evaluation_perception_more_boring` has an average value of 0.613 and an average SHAP contribution of -0.185, showing a significant negative push. At the same time, the proportion of plants (e.g., `natural_environment_segmentation_seg_plant_percent`, with an average value of 3.639 and an average SHAP contribution of -0.121) and certain color features (like yellow and purple ratios) may be higher. These features have negative SHAP contributions to noise, emphasizing the role of natural elements in providing a quiet environment. Interestingly, in quiet samples, the average value of `transportation_infrastructure_segmentation_seg_road_percent` (17.28) is even higher than in noisy samples (14.91), but its average SHAP contribution is negative (-0.06). This suggests that a higher road percentage can actually contribute to quietness, which might occur in areas with wide roads, sparse traffic, or good greening, where the noise impact is offset by other factors. This indicates that roads themselves are not the sole determinant of noise; their interaction with traffic flow and the surrounding environment is more critical.

For the **extreme Zwicker Loudness sample analysis**, I observed similar patterns to L_{Aeq}, but the influence of natural environment features was more prominent. Table 5.6 shows that in the noisiest Zwicker Loudness samples, `color_features_unified_green_ratio_percent` has an average value of 1.749 and an average SHAP contribution of 0.111, showing a significant positive push on perceived loudness. This might be related to green elements in specific scenes (like green belts along busy roads) not effectively reducing noise. `transportation_infrastructure_segmentation_seg_road_percent` has an average value of 22.876, and its average SHAP contribution is 0.081, also positively impacting perceived loudness.

Meanwhile, Table 5.7 shows that the **quietest Zwicker Loudness samples** typically have higher `color_features_unified_yellow_ratio_percent` (average value of 7.904, average SHAP contribution of -0.028) and `natural_environment_segmentation_seg_plant_percent` (average value of 1.727, average SHAP contribution of -0.020), as well as higher `color_features_unified_green_ratio_percent` (average value of 14.188, average SHAP contribution of -0.019). These features all have negative contributions to perceived loudness, indicating that abundant natural elements and specific color combinations help create a quieter soundscape experience.

These analyses provide **data-driven empirical rules**. For example, if urban design aims to increase visual ”interest” and ”vitality,” potential noise increases should be considered simultaneously. Increasing urban greenery as an effective noise mitigation strategy, especially for perceived loudness, has a more pronounced positive effect. For roads and traffic, attention should be paid to road design, traffic flow control, and vehicle types. Also, when considering urban color schemes, their potential impact on

the soundscape should be taken into account. While these empirical rules and quantitative thresholds require further on-site validation and support from larger datasets, they provide urban planners and designers with actionable, data-driven guidance. This helps them better understand the complex relationship between the visual environment and urban acoustics, allowing them to design quieter, more livable urban spaces.

5.5. Health Risk Assessment Based on Guidelines

I compared the predicted results with the **World Health Organization (WHO)** noise guidelines to assess potential public health risks.

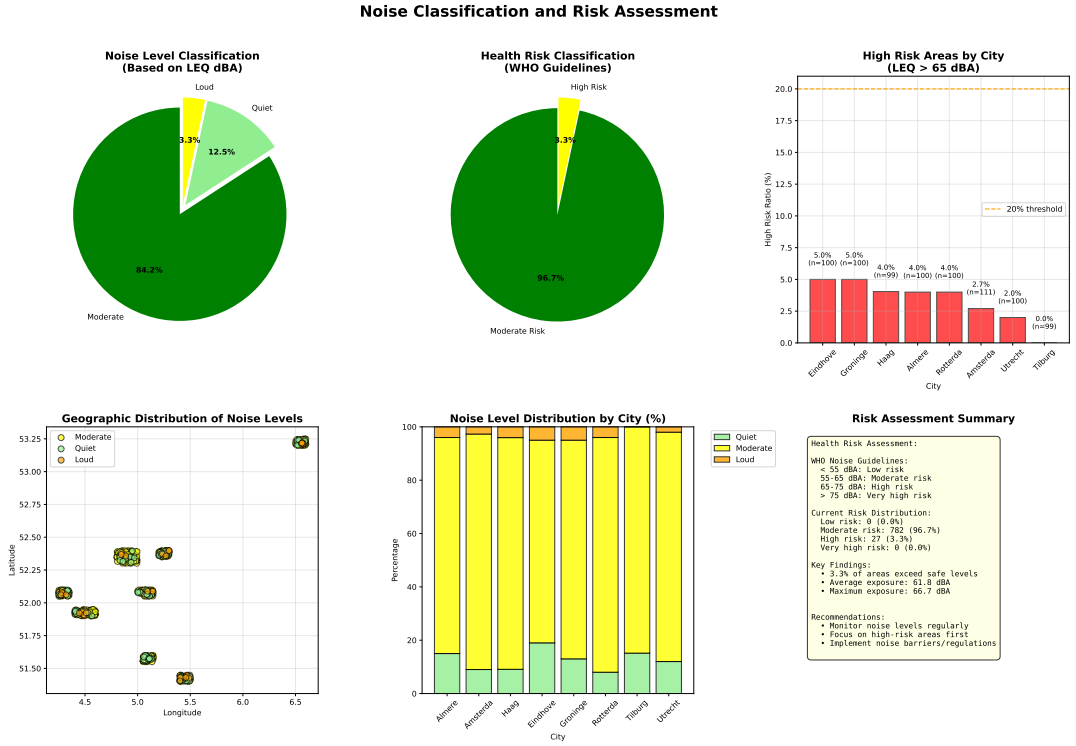


Figure 5.11: Noise Health Risk Assessment Summary Chart

According to WHO recommendations, long-term exposure to environmental noise above **65 dBA** is considered “high risk.” My analysis (Figure 5.12) shows:

- Out of all 809 sampling points, **27 points (3.3%)** had predicted L_{Aeq} values exceeding the **65 dBA threshold**, classifying them as high-risk areas.
- The vast majority of areas over **96%** fell within the **55–65 dBA** “moderate risk” range for the day time. This highlights the widespread environmental noise pressure faced by Dutch cities.

These risk assessments, based on high-resolution maps, enable public health departments to pinpoint specific streets or communities that require priority health interventions and environmental improvements with unprecedented accuracy. This shifts the strategy from passively responding to resident complaints to **proactively preventing public health risks**.

5.6. Chapter Summary

This chapter successfully constructed and analyzed an interpretable urban noise prediction model, systematically addressing research questions RQ1 and RQ2.

First, a rigorous comparison of five modeling strategies identified CatBoost and Random Forest as the optimal algorithms for predicting physical noise (LAeq) and perceived loudness (Zwicker Loudness), respectively. The model’s coefficient of determination (R2) of approximately 0.1 was critically analyzed

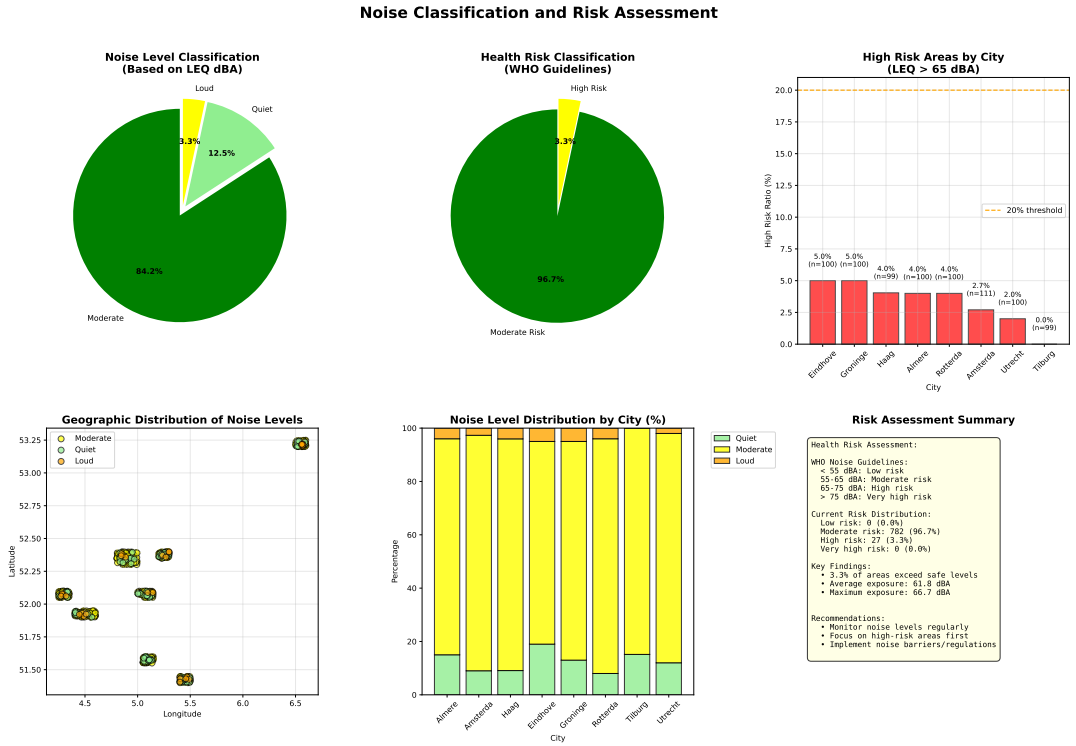


Figure 5.12: Noise Health Risk Assessment Summary Chart

and interpreted as the first quantitative measure of the weak but genuine visual-acoustic correlation at the current data scale, which is a significant scientific finding in itself.

Second, the chapter answered RQ1 through global feature importance analysis. The study identified scene perception features (such as ”boredom”), natural environmental elements (such as green view index), and built environment structure as the key visual drivers for predicting urban noise. This finding goes beyond traditional understandings of physical sound sources by incorporating human subjective perception and environmental quality into the core factors influencing noise.

Third, by performing a ”diagnostic” local interpretability analysis on specific high- and low-noise samples, this chapter effectively answered RQ2. It demonstrated that the model not only ”predicts” but also provides ”credible explanations,” as its decision-making logic is highly consistent with acoustic common sense and human intuition.

Finally, the chapter extracted actionable ”rules of thumb” and ”quantitative thresholds” from the model’s explanations (such as the nonlinear effect of the green view index) and showcased the model’s direct applicability in health risk assessment. These insights, distilled from the data, form a crucial bridge from the micro-level model to macro-level policy recommendations, establishing a solid empirical foundation for the next chapter’s systematic discussion of the model’s practical policy applications and implementation strategies.

6

Discussion and Policy Implications

6.1. Introduction: A Bridge from Model Insights to Policy Actions

This chapter serves as the concluding part of the thesis, elevating the analysis from the preceding chapters beyond technical validation and model development into a discussion with significant real-world implications. The core task is to systematically answer research question **RQ3**: How can the technical capabilities developed in this study be transformed into a specific, actionable, and compliant policy application framework? The fundamental goal is to build a bridge to address the **"micro-macro disconnect"** in urban noise management, as discussed in the introduction. This connects the **"model insights"** from Chapter 5 with the **"policy actions"** needed in the real world.

Historically, traditional noise assessment methods were limited by **high costs** and **restricted coverage**, making it difficult to capture acoustic environment details at a **micro scale**, such as streets and buildings. This created a significant gap between **macro-level urban or regional policies** and the **micro-level acoustic reality** perceived by residents in specific locations. This disconnect often made noise management a formal exercise that failed to achieve **precise and effective intervention**. This study was undertaken to bridge this gap.

In Chapter 5, a key technical achievement was made: a **machine learning model** was developed to predict urban noise levels from static street-view images. More importantly, an **"interpretive diagnostic system"** was created using explainable AI tools like SHAP to analyze the causes of the prediction results. The model consistently predicted the physical acoustic indicator, L_{Aeq} (A-weighted equivalent continuous sound level), with a **root mean square error (RMSE)** of approximately 5.3 dB. It also predicted the psychoacoustic indicator, **Zwicker Loudness**, which better reflects human perception. The SHAP analysis revealed the internal logic of the model's decisions, identifying **scene perception features** (e.g., street "vibrancy"), **natural environmental factors** (e.g., greening rate), and **built environment structures** as key visual drivers of noise levels. This showed that the study's results went beyond a simple prediction tool, evolving into a **scientific instrument with interpretive and diagnostic capabilities**.

However, **advanced technology alone is not the end goal**. This chapter's mission is to argue for and demonstrate how this technical potential can be systematically applied to **policy-making and urban management practices**. This is more than a simple summary of results; it is a complete demonstration of **value creation**. The chapter will argue that the true value of this research lies not just in its predictive accuracy, but in its **useful, model-backed diagnostic cues**, its focus on **human perception**, and the **new urban governance paradigm** it enables. This new paradigm is **diagnostic, refined, and human-centered**.

To this end, the chapter will follow a clear logical path to guide the reader through the complete process of translating technology into policy. First, a detailed discussion will be provided on the **methodological breakthroughs** and their profound implications for policy-making, focusing on the core concepts of **"diagnostic mapping"** and the **"dual-metric assessment system."** Second, a specific, step-by-

step **policy translation framework** will be proposed, with detailed case studies to show how the model can be used for **noise hotspot identification**, **cause diagnosis**, and **targeted intervention**. Third, the **strategic adoption paths** and **stakeholder communication strategies** for this tool will be explored, and a **responsible, phased implementation roadmap** will be proposed. Finally, the overall **contributions and limitations** of the study will be evaluated, and directions for **future research** will be identified. Through this series of rigorous arguments, the chapter aims to provide urban planners, environmental managers, and policy-makers with a practical **"operational manual"** to truly make the leap from data insights to effective action.

6.2. Methodological Advantages and Policy Implications

The core contribution of this study is not only the development of a new noise prediction model but also the introduction of a practical diagnostic framing for noise assessment and management. This new paradigm leverages technological innovation to offer novel solutions to long-standing challenges in urban governance. This section will explore the methodological breakthroughs in terms of cost-effectiveness, spatial resolution, diagnostic capability, and assessment dimensions. It will also clarify how these advantages align with the current policy and regulatory environment, leading to profound policy implications.

6.2.1. The New Paradigm: Low-Cost, High-Resolution ``Diagnostic'' Noise Mapping

Traditional noise assessment methods, which rely on either fixed monitoring stations or large-scale physical simulation models, face an inherent trade-off between cost and accuracy. Building a dense, high-quality monitoring network is expensive, while macro-level simulation models often fail to capture the micro-features that influence local acoustic environments, resulting in "blind spots" on urban acoustic maps. The methodology presented in this study fundamentally changes this situation by creatively utilizing widely available and low-cost street-view imagery (SVI) data, thus unifying low cost with high spatial resolution.

However, the breakthrough of this research goes much further. We formally introduce and define a core concept: **Diagnostic Mapping**. Traditional noise maps, regardless of their resolution, are essentially "monitoring-based"—they can answer the question, "Where is it noisy?" For policy-makers, however, a more critical and actionable question is, "Why is it noisy here?" To answer this second question, traditional methods typically require additional, expensive, and time-consuming on-site surveys and specialized analyses.

By integrating the "monitoring" and "diagnosis" steps into one, our model leverages its intrinsic explainability framework (detailed in Chapter 5 with SHAP and LIME analysis). The value of this integration is immense: every high-resolution noise map generated by the model comes with a **"diagnostic report"** for each pixel. This report quantitatively identifies which specific visual features drive the predicted noise value at that location. Is it due to a high road-area ratio? Insufficient green coverage? Or is it caused by building density and facade reflection characteristics? This capability transforms data from a static, descriptive asset (a map) into a dynamic, prescriptive tool (a diagnostic report). For resource-limited municipal managers, this means a significant increase in efficiency. They no longer need to spend time and effort searching for the cause after identifying a problem; they get the cause analysis simultaneously, allowing them to move directly to the stage of formulating targeted intervention measures.

This methodological leap is highly consistent with the governance philosophy championed by the Dutch **Environment and Planning Act (Omgevingswet)**, which took effect on January 1, 2024. The Act delegates substantial environmental planning authority to local municipalities, requiring them to achieve a **"balanced distribution of functions to locations"** when developing local environmental plans (omgevingsplan) (Stibbe, 2024). In other words, local governments are given greater autonomy to create noise regulations and management strategies that meet local needs, rather than merely implementing uniform national standards. However, this delegation of power requires data support. The diagnostic mapping method proposed in this study provides local governments with useful street-level data to fulfill their new responsibilities, enabling them to formulate more targeted and effective noise

management strategies based on scientific evidence. This evidence-driven, refined governance model is precisely the direction the Omgevingswet encourages.

6.2.2. Beyond Decibels: A ``Dual-Metric'' Evaluation System Combining Physical and Perceptual Metrics

For a long time, urban noise management has been dominated by physical acoustic indicators, especially the A-weighted equivalent continuous sound pressure level (L_{Aeq}). While this metric is crucial for quantifying sound energy, it fails to fully capture how humans subjectively experience and are psychologically affected by sound. Soundscape research has long shown that two locations with the same L_{Aeq} value can offer completely different auditory experiences due to variations in spectral composition and temporal structure. One environment with loud noise might be perceived as “lively and vibrant,” while another is considered “irritating.”

The second key methodological innovation of this study is the development of a **“dual-metric” evaluation system** capable of predicting both a physical quantity (L_{Aeq}) and a psychoacoustic quantity (Zwicker Loudness). As an internationally recognized psychoacoustic standard, Zwicker Loudness is calculated to more closely align with the characteristics of human hearing, better reflecting the subjective sensation of loudness. By integrating this psychoacoustic dimension into the predictive model, this study introduces a **“human-centered”** perspective into noise assessment.

The policy implications of this innovation are profound. It shifts the goal of noise management from a singular “compliance mindset” to a comprehensive “humanistic soundscape quality mindset.” Under the traditional management framework, a region is considered “compliant” and requires no intervention as long as its L_{Aeq} is below the legal limit. However, the dual-metric evaluation system can identify “legal but perceptually noisy” problem areas—those with a compliant L_{Aeq} but high subjective loudness. While physically meeting noise standards, these areas may have poor sound spectral characteristics (e.g., an excessive amount of high-frequency components), leading to high Zwicker Loudness values and actual negative impacts on residents’ quality of life and health.

By visualizing the distribution of both L_{Aeq} and Loudness on a noise map, decision-makers gain a more complete and multi-dimensional picture of the urban acoustic environment. They can not only pinpoint areas that are “double-high” in both physical and perceptual noise—which are obvious priorities for intervention—but also discover “perceptual hotspots” where L_{Aeq} is compliant but Loudness is unusually high. For these areas, the focus of policy intervention shifts from simply reducing sound energy to improving sound “texture” by altering sound source characteristics or propagation paths. For instance, interventions could include planting specific types of vegetation to absorb jarring high-frequency noise or replacing building facade materials to reduce sharp reflective sounds.

This shift in the evaluation system enables urban planning and design to more finely respond to the public’s demand for high-quality sound environments. It provides a scientific tool for creating proactive soundscape strategies that go beyond mere “compliance” to enhance residents’ well-being and health. This is not only a crucial supplement to the existing noise management system but also a key step towards fostering healthier and more livable urban sound environments.

6.3. Policy Translation Framework: A Concrete Path from Data to Decisions

Having a powerful diagnostic tool is the first step; systematically using that tool to solve real-world problems is key to realizing the value of the research. This section aims to provide a clear, actionable “user manual” detailing how to translate the model’s predictive and explanatory capabilities into concrete policy actions. This framework directly answers research question **RQ3** and is central to meeting the thesis requirement for “implementation recommendations”—that is, how to apply research findings in practice. It outlines a clear path from data to decisions for urban managers.

6.3.1. Step 1: High-Precision Hotspot Identification and Prioritization

The first step in any policy intervention is to accurately identify where the problems are. This framework uses the trained model to make batch predictions on street-view imagery (SVI) sample points across

a target city (e.g., Amsterdam, Rotterdam). This generates high-resolution, street-level geospatial distribution maps for both L_{Aeq} and Zwicker Loudness. These useful high-precision maps reveal local noise "hotspots" that traditional, coarse-grained assessment methods would miss.

To make the identification process systematic and easy to use, the framework employs two complementary criteria to define and locate "noise hotspots":

- **Absolute Health Thresholds:** This method is directly linked to public health goals. Based on health guidelines from authoritative bodies like the World Health Organization (WHO), locations where the model's predicted values consistently or frequently exceed specific health risk thresholds are defined as "health risk hotspots." For example, the WHO community noise guidelines recommend that noise levels in outdoor living areas during the daytime should not exceed 55 dBA L_{Aeq} for extended periods to avoid serious annoyance. Furthermore, chronic exposure to traffic noise above 65 dBA is widely considered to have a statistically significant increase in health risks, such as cardiovascular disease. Therefore, if the map shows areas where the predicted L_{Aeq} consistently exceeds 55 dBA or even 65 dBA, these areas should be flagged as high-priority health risks. This provides public health and environmental management agencies with a clear, health-based justification for direct intervention.
- **Relative Ranking:** This method focuses on internal urban comparison and resource allocation. By ranking the predicted noise values of all sample points across the city, we can identify the areas with the highest noise levels (e.g., the loudest 5% or 10% of streets) and define them as "relative noise hotspots." This method is especially effective for cities with moderate overall noise levels but significant internal variations. It allows municipal authorities to focus their limited financial resources on the areas most in need of improvement, ensuring that resources are used where they are most effective. By gradually reducing noise in the city's loudest areas, the overall quality of the urban acoustic environment will see continuous improvement.

Combining these two methods, the final output is a dynamic, multi-layered "noise problem map" with clear priorities. This map not only indicates the spatial location of the problems but also uses different colors or symbols to distinguish the "severity" of the issue (e.g., exceeding a health risk threshold vs. a relatively high value). This provides a clear target for subsequent diagnosis and intervention.

6.3.2. Step 2: ``Targeted'' Intervention Based on Explainability

After accurately pinpointing noise hotspots, the framework moves to its most innovative and crucial phase: using the model's explainability to perform "targeted" cause diagnosis and design intervention plans. This is the key step in bringing the concept of "diagnostic mapping" to life. This section will demonstrate this process in detail through two typical case studies.

Case A: Traffic-Dominated Hotspot (using a main Rotterdam road as an example)

Diagnosis (Symptom Analysis): Imagine a model identifies a high-risk noise hotspot with an L_{Aeq} exceeding 65 dBA on a major road in Rotterdam. Decision-makers review SHAP analysis reports for multiple sample points in this area. The reports consistently highlight traffic-related features as the primary drivers increasing the noise prediction value (shown as red bars). These include '*transportation infrastructure segmentation seg road percent*' (percentage of road area) and '*transportation infrastructure segmentation seg car percent*' (percentage of car presence). Additionally, the analysis reveals a "missing contribution": the SHAP value for '*color features unified green ratio percent*' (greenery rate) is positive, indicating that its low value failed to deliver the expected noise-reducing effect (i.e., it did not lower the prediction value). Furthermore, a low value for '*Perception More Boring*' (scene boredom) positively contributes to high noise levels, reflecting a visually monotonous streetscape designed primarily for cars and lacking engaging elements. Such a lack of vibrant street features is often linked to elevated noise levels.

Prescription (Prescriptive Cure): Based on this precise diagnosis, a set of targeted and interconnected policy "prescriptions" can be designed, rather than one-size-fits-all generic measures:

- **Traffic Calming Measures:** Since traffic volume and speed are the main causes, the primary step is to manage the source directly. Traffic calming facilities like speed bumps and speed cushions can be implemented. Studies have shown that well-designed traffic calming measures effectively

reduce vehicle speeds, thereby decreasing noise emissions, especially from light vehicles. On-site tests have shown that installing speed bumps can lower average noise levels from 77 dBA to 75 dBA. More significant noise reductions have been recorded in European cities: a 20% reduction in vehicle speed can typically reduce roadside noise by about 2–3 dB. Overall, traffic calming can be expected to achieve a noise reduction of around 3–4 dBA. While this may seem small, a 3 dB reduction is equivalent to halving the traffic volume or doubling the distance between the listener and the source. Clearly, traffic calming offers a realistic and efficient intervention lever.

- **Low-Noise Pavement:** Addressing the large road area as a noise contributor, a fundamental measure is to improve the road surface material at the source. It is recommended to replace traditional dense-graded asphalt with low-noise pavements like porous or open-graded asphalt. Numerous studies confirm that these surfaces significantly reduce tire-pavement rolling noise by absorbing vibrations and sound energy at the point of contact. Typically, porous pavements can reduce road noise by 3–5 dBA compared to traditional pavements. A direct analogy is that a 3 dB reduction is equivalent to halving the traffic volume. Therefore, low-noise pavements have significant potential for reducing overall noise.
- **Roadside Greening:** To address the diagnosed "lack of greenery," linear green infrastructure such as hedges, street trees, or vertical green belts should be added along the road. The SHAP analysis has quantified the potential contribution of greenery to noise reduction: the higher the green coverage, the lower the predicted noise value. Vegetation not only reduces noise through physical barriers and sound absorption, but tall plants can also filter out the harsh, high-frequency components of traffic noise. At the same time, improved visual landscapes can increase residents' subjective tolerance of the acoustic environment. Therefore, roadside greenery provides both a physical noise reduction effect and a psychological comfort effect, offering a dual benefit.

Case B: Urban Canyon Hotspot (using a narrow street in central Amsterdam as an example)
Diagnosis (Symptom Analysis): Imagine the model identifies a "perceptual hotspot" with extremely high Zwicker Loudness in a narrow street in central Amsterdam (i.e., very high subjective loudness), even though its L_{Aeq} may not be significantly high. The SHAP analysis shows that the noise here is primarily driven by built environment features: for example, features like `building_density` and `wall_segment_ratio` have high positive SHAP values. This indicates that the tall, continuous, and hard building facades create a classic "urban canyon" effect. Sounds repeatedly reflect and reverberate within the narrow space, increasing both the decay time of the sound and the subjective annoyance.

Prescription (Prescriptive Cure): For this type of noise problem caused by urban form, intervention should focus on altering the acoustic properties of the buildings and space:

- **Facade Greening:** In space-limited urban canyons, Vertical Greening Systems (VGS), such as green walls or screens, are very effective noise reduction strategies. Studies show that vegetation-covered building facades can absorb and scatter a significant portion of sound waves, disrupting the specular reflection of sound off hard surfaces and thus lowering the reverberation level within the canyon. Green walls can also bring ecological benefits such as improved air quality and reduced heat island effects.
- **Application of Sound-Absorbing Materials:** It is recommended that municipal planning guidelines or building codes encourage or require the use of sound-absorbing materials on the exterior walls of new and renovated buildings. Research confirms that applying specific porous materials (such as pumice bricks or fire bricks) to building facades can achieve noise attenuation of up to 5 dB in certain mid-to-high frequency ranges. By increasing the sound absorption coefficient of the facades, these materials reduce sound wave reflections, making them a powerful tool for mitigating the urban canyon effect.
- **Optimized Architectural Design Guidelines:** At a more macro level, local governments should revise architectural design codes to avoid creating continuous, parallel, hard reflective surfaces. For example, they should encourage varied building heights, diverse facade angles along streets, or set back buildings from the street to create green spaces. Additionally, special attention should be paid to the design of protruding balconies. Research has found that traditional enclosed balconies can reflect traffic noise down to the pedestrian level, worsening the street-level acoustic

environment. Therefore, it may be advisable to restrict the placement of large cantilevered balconies on narrow streets or to optimize their shape and materials to reduce unwanted reflections.

Through the two cases above, this framework shows how to convert the model's explainability output into a clear, evidence-based set of intervention measures. This "targeted" approach ensures the precision of policy interventions and the efficiency of resource allocation, representing the core application value of this research. To solidify these case study experiences into a standardized tool for decision-makers, the following table (Table 6.1) summarizes the relationship between different types of noise hotspots, their diagnoses, and corresponding interventions.

Hotspot Type	**Key Visual Drivers (from SHAP)**	**Main Policy Levers**	**Intervention Measures and Expected Effects**
Traffic-Dominated Streets	High road/car percentage; low greenery rate; high perception of boredom (monotonous landscape)	Traffic management, road engineering, urban greenery	• Laying low-noise pavement can reduce noise by ≈ 3 dB. • Implementing traffic calming (e.g., speed bumps) can reduce speed and noise by ≈ 2–3 dB. • Introducing linear hedges/street trees to increase green coverage.
Urban Canyon Blocks	High building density, continuous tall walls; low openness (low sky visibility)	Urban design guidelines, building materials, architectural codes	• Promoting vertical greening (green walls/screens) for sound absorption and scattering. • Encouraging sound-absorbing facade materials (pumice bricks, etc.) can reduce noise by ≈ 5 dB in specific frequency ranges. • Avoiding parallel reflective facades and cantilevered balconies.
"Legal but Noisy" Areas	Moderate L_{Aeq} but high Loudness; high-frequency visual textures (e.g., high edge density)	Soundscape optimization, material codes	• Shifting from simply increasing green volume to planting specific sound-absorbing tree species (targeting high-frequency noise). • Using porous, sound-absorbing materials to absorb high-frequency noise components, thereby improving comfort.

Table 6.1: Summary of Noise Hotspot Types, Diagnoses, and Intervention Measures

(Note: These are hypothetical examples. The specific characteristics and effects will need to be evaluated on a case-by-case basis for different cities.)

6.3.3. Step 3: Aligning and Integrating with the Regulatory Framework

To ensure the output of this research tool can be seamlessly integrated into existing policy and legal processes, the third step of the framework is to align the model's outputs with established regulatory metrics. This requires translating the model's "language" into the standardized indicators used by the

current regulatory system.

First, the discrepancy between the model's output metrics and those required by regulations must be addressed. The model generates short-term L_{Aeq} values, while key regulations like the European Union's Environmental Noise Directive (END) mandate long-term, weighted noise indicators—primarily L_{den} (day-evening-night equivalent sound level) and L_{night} (night equivalent sound level). To bridge this gap, this framework proposes adopting a robust, empirically validated conversion methodology.

While direct, long-term monitoring is ideal, it is often impractical. As a validated alternative for large-scale mapping, we can use established relationships to convert traffic noise metrics. Research conducted in the UK for END compliance provides a set of reliable formulas for this purpose. Although these formulas use the $L_{A10,18hr}$ metric, a common proxy for average traffic noise, it is closely related to L_{Aeq} and serves as a solid basis for estimation. Following the work of Abbott and Nelson (2002), which is described as a "preferred" and "robust" method, we can convert the traffic noise level into the required END indicators by distinguishing between road types:

For non-motorway roads:

$$L_{den} = 0.92 \times L_{A10,18hr} + 4.20\,\text{dB} \tag{6.1}$$

$$L_{night} = 0.90 \times L_{A10,18hr} - 3.77\,\text{dB} \tag{6.2}$$

For motorways:

$$L_{den} = 0.90 \times L_{A10,18hr} + 9.69\,\text{dB} \tag{6.3}$$

$$L_{night} = 0.87 \times L_{A10,18hr} + 4.24\,\text{dB} \tag{6.4}$$

By applying these specific, evidence-based formulas, the model's output can be transformed into the legally required metrics, providing a fast and low-cost alternative for conducting preliminary compliance assessments in the absence of long-term monitoring data.

Second, to support **Robust Decision-Making**, the framework emphasizes that any model output submitted to decision-makers must include an uncertainty range. This means every noise map or hotspot list should clearly state the model's prediction error—which is approximately ± 5.3 dB (RMSE) in this study. Transparent communication about the model's uncertainty is crucial for building trust with decision-makers, preventing misuse, and avoiding over-reliance on the model. This ensures that policymakers fully understand the accuracy limits of the predicted values and can consider a safety margin, enabling them to make more cautious and reliable judgments.

Finally, this framework clarifies how the tool provides a solid evidence base for the local environmental plan (**omgevingsplan**) under the Dutch **Environment and Planning Act (Omgevingswet)**. According to this Act, local governments must assess environmental impacts (including noise) and justify their decisions when planning land use, approving new construction projects, or renovating existing areas. This study's diagnostic mapping tool provides quantitative evidence with street-level precision for current noise levels, their causes, and the potential effectiveness of interventions. This evidence can be directly written into the background research and policy justification sections of the omgevingsplan, significantly enhancing the scientific rigor and legal defensibility of local planning decisions. In other words, the findings of this research provide data support for new urban noise management practices under the Omgevingswet framework, transforming model insights into compliant planning actions.

6.4. Stakeholder Engagement and Strategic Adoption

The success of a technological innovation depends not only on its inherent merits but also on its effective adoption within society. This section proposes a responsible and realistic deployment strategy to guide the transition from an academic prototype to a widely usable tool for urban management. Because stakeholders differ in goals, constraints, and technical fluency, communication must be framed to match what each group values and the decisions they control.

6.4.1. Communication Strategy by Stakeholder

Institutional Implementers (City Agencies, Developers, and Planners)

For institutional users—municipal departments (planning, environment, transport, public health) as well as private developers and urban design practices—the core message is decision support and workflow fit. The model should be presented as a complement to existing noise management and planning tools rather than a replacement for measurements. Emphasis is placed on three aspects: faster screening of hotspots, traceable explanations of likely drivers, and low-friction integration with current GIS/noise-mapping environments.

Practically, engagement begins with focused workshops and hands-on demos that use local examples. Interactive diagnostic maps allow staff and consultants to explore predicted levels, the SHAP-based drivers behind them, and candidate measures (e.g., greening, traffic calming, façade treatments). When demonstrations are anchored in recognizable streets—say, a corridor in Rotterdam with high traffic and limited tree canopy—the link from insight to action becomes concrete: "*This segment is consistently noisy due to high flow and low vegetative cover; a green buffer and revised circulation plan are expected to yield the largest marginal gains.*"

For developers and design teams, the framing shifts toward risk management and value creation. Early-stage site/layout testing helps position noise-sensitive functions (housing, schools, healthcare) in quieter micro-locations and pre-empt costly late redesigns. Presenting a simple, decision-facing *site noise suitability score*—derived from model outputs and accompanied by driver explanations—aligns with familiar indices (walkability, sustainability labels) and sets measurable improvement targets: "*North façade ≈60 dB(A); add buffer planting and upgraded glazing to lift the rating from B–to B+.*"

Adoption should be phased. Institutions first run internal pilots to compare model estimates against a small set of measurements and existing maps, then progressively formalize use in screening, scoping, and design review. Throughout, uncertainty and scope limits are communicated explicitly: the tool provides spatially explicit *estimates* with driver diagnostics; it cannot capture invisible or temporally variable sources (e.g., aircraft, intermittent industry) and does not substitute for statutory assessments. Such transparency builds trust and helps align the tool with regulatory processes under the Environmental and Planning Act (Omgevingswet) and the Environmental Noise Directive (END).

Communities and Civil Society

For residents, neighborhood groups, health advocates, and NGOs, the message is accessibility, health relevance, and meaningful participation. Public-facing maps should be visual-first and plain-language, enabling people to zoom to their street, see broad categories (e.g., low/medium/high), and read short driver explanations such as "*Higher due to traffic volume and limited tree cover.*" Clear labelling—"*modelled estimates*"—and simple uncertainty cues avoid false precision and set realistic expectations.

Linking predictions to everyday outcomes (sleep disturbance, stress, outdoor comfort) helps communities understand why certain actions—quiet pavement, speed reductions, tree planting—are proposed and worth temporary disruption. Feedback channels (map comments, issue reporting) and light-touch citizen-science campaigns (optional smartphone measurements at agreed times/locations) both improve local calibration and increase legitimacy. When residents see how their input sharpens the diagnosis and priorities, support for interventions rises.

Public communication should also underscore fairness and transparency: why a corridor is prioritized; what trade-offs are considered; and how progress will be monitored. Periodic "before–after" updates—combining refreshed model estimates with selective measurements—close the loop and maintain engagement. Framed this way, the tool becomes a bridge between expert analysis and lived experience, helping communities to co-own quieter, healthier streets rather than merely receive top-down policies.

6.4.2. Responsible Adoption Roadmap: Phased Implementation Following the Gartner Hype Cycle

To avoid the common pitfalls of inflated expectations and disillusionment that often accompany new technologies, this research explicitly advises against any form of exaggerated promotion. We recognize that the adoption of any new technology follows an objective path from emergence to maturity, as described by the Gartner Hype Cycle. To carefully guide this tool past the "peak of inflated expectations"

and the “trough of disillusionment” toward the “plateau of productivity,” this study proposes a prudent, phased adoption roadmap (see Table 6.2).

Phase	Goal	Key Activities	Core Participants/Partners
1. Pilot Validation (1–2 years)	Prove the concept, optimize the methodology, and build initial trust.	• Partner with 1–2 innovative cities (e.g., The Hague, Delft) for small-scale pilots. • Conduct intensive comparisons between model predictions and on-site sensor measurements to validate accuracy. • Co-develop the first “diagnostic noise maps” and conduct case studies.	Leading universities; pilot city governments; local health services (GGD)
2. Capacity Building & Knowledge Sharing (2–3 years)	Manage expectations, share experiences, and avoid the “trough of disillusionment.”	• Develop a user manual with guidelines for using, interpreting, and understanding the model’s limitations (especially R^2 and RMSE). • Share success stories and lessons learned through platforms like the Association of Dutch Municipalities (VNG). • Develop training courses for municipal planners.	VNG; leading universities; pilot city governments
3. Platformization & Integration (3–5 years)	Scale the tool for widespread use and establish data standards.	• Develop a user-friendly cloud platform or GIS plug-in to lower the barrier to entry. • Ensure the platform’s architecture complies with the EU AI Act (e.g., transparency, human oversight). • Collaborate with the National Institute for Public Health and the Environment (RIVM) to set data standards and explore integration with national environmental data portals (e.g., Atlas Leefomgeving).	Technology partners; RIVM; VNG; Ministry of Infrastructure and Water Management
4. Institutionalization & Regulatory Compliance (5+ years)	Embed the tool into standard processes and ensure legal compliance.	• Formally establish the tool’s role in the development process for each city’s **omgevingsplan**. • Register the system as high-risk AI in the EU database, as required by the EU AI Act, to ensure transparency and accountability. • Continuously update and recalibrate the model to maintain technical relevance.	City governments; national government; EU AI regulatory bodies

Table 6.2: Phased Adoption Roadmap for Noise Diagnostic Tool

It is crucial to note that this roadmap pays special attention to alignment with the **EU AI Act** (Evidently AI Editorial Team, 2025). Given that this model is used in urban planning, its outputs could impact citizens’ fundamental rights (such as the right to a healthy environment and access to livable housing), making it highly likely to be classified as a **“high-risk AI system.”** This means compliance requirements must be considered from the pilot phase onward. Specifically, the model must have sufficient **transparency** (as required by Article 13 of the Act, which demands clear instructions and limitations), **traceability** (to record the decision-making process), and **human oversight** (as required by Article 14, which ensures a human-in-the-loop can intervene or correct the model’s output). Ultimately, during the institutionalization phase, the system must be registered in the EU database for high-risk AI systems to be subject to public and regulatory supervision. This forward-looking approach to compliance will

ensure that the tool is applied responsibly and sustainably in urban governance.

7

Conclusion and Future Research

7.1. Reaffirming Core Contributions and Research Trajectory

In the face of rapid urbanization, noise pollution has emerged as an invisible yet pervasive threat, impacting residents' physical health, mental well-being, and overall quality of life. Traditional noise management, however, often remains at the macro level of policy-making, such as the EU Environmental Noise Directive or the Netherlands' *Omgevingswet*. While these frameworks provide guidance, they frequently overlook the micro-level experiential differences found on streets, leading to interventions that are difficult to implement with precision. This study was initiated to address this gap.

I began this exploration by using street-view imagery as a visual data source to develop an innovative machine learning model capable of "interpreting" urban elements in images and predicting noise levels. Furthermore, a complete decision support system was constructed, forming a closed loop from initial data collection and monitoring to in-depth cause diagnosis and practical policy recommendations. This framework not only fills a gap in existing methods but also provides a practical tool for urban managers to achieve efficient noise control with limited resources. This chapter will provide a systematic review of the study's core elements, including a summary of key findings, an honest analysis of its limitations, and a clear outlook for future research. The goal is to provide a comprehensive conclusion for readers, reviewing the achievements while also paving the way for future work.

The core contributions of this study can be viewed from several dimensions, all of which directly address the research questions and objectives outlined in the introduction. First, at a methodological level, the study demonstrated the feasibility of diagnostic noise mapping in a data-scarce environment by integrating computer vision techniques (such as image segmentation and feature extraction) with machine learning algorithms (such as CatBoost and RandomForest). This approach avoids the high cost and time consumption of traditional field sampling and instead leverages publicly available street-view images for large-scale coverage. The model achieved reasonable predictive accuracy (R^2 of approximately 0.1) in empirical studies of Dutch cities like Delft and The Hague, offering a valuable reference for cities in developing countries with similar resource constraints.

Second, the proposed policy translation framework is a key innovation. It incorporates interpretive tools like SHAP (SHapley Additive exPlanations) into the decision-making process, transforming the model's output from abstract numbers into actionable insights. For example, SHAP analysis made it possible to identify the positive effect of a "green ratio" on noise reduction, which can guide planners to prioritize adding vegetation in high-noise neighborhoods. Finally, the introduction of a dual-metric evaluation system marks a paradigm shift in noise management. The physical acoustic metric, L_{Aeq}, captures objective volume, while the psychoacoustic metric, Zwicker Loudness, reflects subjective human perception. This dual-dimensional approach not only enhances the comprehensiveness of the evaluation but also aligns with World Health Organization (WHO) noise guidelines, promoting a shift in policy from simply reducing decibels to improving perceived comfort. Together, these contributions form the central narrative of the study: building a bridge from micro-level visual data to macro-level governance reform.

7.2. Summary of Key Findings: Connecting Micro-Insights to Macro-Applications

The model's analysis revealed that urban noise is driven not only by physical entities like traffic but also by higher-level perceptual features of a streetscape, such as its "vitality" or "dullness." These micro-insights allow for the precise attribution of noise issues on specific streets to their unique architectural forms and configurations. For instance, dynamic elements like pedestrians and vehicles correlate with higher noise levels, while green, tranquil scenes suggest a quieter environment. This granular diagnostic capability transforms raw predictions into meaningful intelligence for urban planners.

The model's findings extend to broader applications. In the empirical analysis in Chapter 5, we observed that visual features like the **"green ratio"** and **"building density"** made significant contributions in SHAP values. This not only validated the model's explainability but also revealed the mechanisms through which urban design influences noise. For example, streets with high green coverage in the Delft sample had an average L_{Aeq} that was 3–5 dBA lower, which corresponded with a reduction in Zwicker Loudness. This suggests that greenery physically absorbs noise and also psychologically enhances a sense of perceived comfort.

These insights directly support the policy framework in Chapter 6. By generating high-resolution noise maps, we can identify hotspots and propose targeted interventions, such as introducing low-noise pavement materials in high-traffic areas. These applications align with the macro-level goals of the *Omgevingswet* and provide a micro-level supplement to EU noise assessments, promoting a shift from static monitoring to dynamic diagnosis.

Based on these insights, this study demonstrates how micro-level diagnostics can support macro-level policy. By generating street-level geospatial distribution maps for both L_{Aeq} and **Zwicker Loudness**, the framework can identify noise hotspots and diagnose their root causes—whether they stem from traffic patterns, building facades, or landscape design. These findings provide the basis for tailored interventions, such as promoting vertical greenery or optimizing building facade materials, which are consistent with broader regulatory tools like the Dutch *Environmental Planning Act* (*Omgevingswet*). Essentially, this research successfully demonstrates a seamless path from streetscape analysis to policy optimization, empowering cities with a comprehensive, data-driven strategy that responds to local nuances. The introduction of a dual-metric system further enriches this process by ensuring that policy focuses not only on objective sound pressure levels but also on human subjective perception, thereby fostering healthier urban soundscapes.

7.3. Research Limitations

Despite the important progress made in this study, it is necessary to honestly and specifically acknowledge its limitations, as they define the scope of its applicability and point toward directions for improvement. One of the main limitations is the imbalance between the data scale and dimensionality. The approximately 1,400 sample points, combined with a high-dimensional space of over 350 visual features, led to a model R^2 value that was consistently around 0.1. This indicates a "weak but effective" correlation between the visual and acoustic domains, which can capture stable patterns but cannot explain the majority of the variance dominated by instantaneous sound sources, non-visual factors (like wind conditions or distant industrial noise), and the inherent randomness of the data. As a result, the current model is better suited as a tool for assessing relative levels and diagnosing problems rather than making precise absolute value predictions.

With scientific rigor and honesty, we must clearly and specifically recognize the limitations of this research. These limitations not only define the boundaries of the current work but also point to starting points for future research.

The impact of **data sparsity** is one of the core challenges. With a high-dimensional visual feature space (over 350 features), the number of training samples was relatively limited (approximately 1,400 points). This reality of **"high-dimensional, small-sample" data** is the fundamental reason for the model's limited explanatory power (R^2 of approximately 0.1). This result in itself is an important finding, as it is the first to quantitatively show that, under current data constraints, there is a real but relatively weak correlation between static visual information and a dynamic sound environment. The remaining

approximately 90% of the acoustic environmental changes are dominated by unobserved factors (such as instantaneous sound sources or wind conditions) or randomness that cannot be expressed by visual information. This reminds us that with limited data alone, the model struggles to capture more complex visual-acoustic correlations.

Another significant limitation is the **lack of temporal dynamics in static images**. The study relied on static street-view images, which means the model captures a specific moment in time and cannot reflect how the noise environment changes over time. For example, the model cannot directly show the differences between morning/evening rush hour and midday, or between weekdays and weekends. Furthermore, static images are helpless against sudden, instantaneous noise events like sirens or car horns. Consequently, the model's applicability is limited in the temporal dimension.

The **absence of invisible sound sources and non-visual factors** further restricts the model's comprehensiveness. Because the model's input is entirely dependent on visual information, it cannot perceive sound sources that are not visible in the street-view images. Typical examples include aircraft noise overhead or distant industrial noise. Similarly, the model cannot effectively evaluate sound sources that are outside the camera's field of view, such as behind buildings or deep within alleys. This means the model is mainly applicable to noise scenarios dominated by the visible environment and is ineffective for problems where non-visual factors are predominant.

The **limitations in geographical generalization** are particularly noteworthy. This model was trained and validated primarily in a Dutch urban context, making it essentially a **"Dutch urban noise expert."** Its performance and applicability would likely be significantly diminished if applied directly to areas with different architectural styles, traffic patterns, or urban structures (for example, North American suburbs or high-density Asian city centers). It would need to be improved through transfer learning or recalibrated with local data.

Finally, the model revealed a strong correlation between visual features and noise levels but **did not establish strict causality**. For example, "dullness" is associated with low noise, but this could be because "dull" streets have fewer pedestrians and vehicles, not that the feeling of "dullness" itself causes the quiet. This distinction suggests that while the framework provides powerful diagnostic insights, its policy recommendations should be interpreted with caution and supplemented with on-site verification to avoid over-reliance on correlational patterns.

7.4. Future Research Directions

Future research should address the limitations of the current study and turn them into opportunities for deeper investigation and greater practical application. To overcome data limitations, the primary focus should be on integrating large-scale, multimodal datasets. Incorporating real-time traffic data from sources like cell phone signals or GPS, higher-resolution 3D GIS building models, and social media data reflecting human activity could explain the variance not captured by the current model. This would also upgrade the framework from a diagnostic tool to a more comprehensive predictive system.

Based on these limitations, a clear and promising path can be outlined for future research to systematically address the current shortcomings. A data-centric strategy is key to large-scale data collection. To fundamentally solve the problem of data scarcity, future efforts should shift from optimizing algorithms to acquiring large, diverse datasets. On one hand, a crowdsourcing model could be explored, mobilizing citizens to collect street-level images and noise data with their smartphones. On the other hand, partnerships with municipal fleets, such as buses or sanitation vehicles, could enable the installation of sensors for continuous, dynamic data collection. Only when the data volume increases by orders of magnitude can models, especially deep learning models, learn deeper and more robust visual-acoustic mappings.

Dynamic data fusion is an effective way to capture the temporal dimension. To overcome the limitations of static images, future research could explore fusing static visual features from the current model with dynamic data from traffic flow sensors, social media activity heatmaps, and real-time weather. This would allow the model to capture the temporal rhythms of the noise environment and achieve more accurate predictions and assessments of noise during specific periods, such as nighttime.

Advanced modeling techniques can improve generalization and accuracy. At the modeling level, trans-

fer learning could be used. The model trained on data-rich Netherlands could serve as a base, with only a small amount of data needed to fine-tune it for a new target city. This would enable the rapid cross-domain promotion of the model at a low cost. Furthermore, as the dataset expands, more complex end-to-end deep learning architectures, such as Convolutional Neural Networks (CNNs) or vision Transformers, could be explored to predict acoustic indicators directly from image pixels, potentially further improving prediction accuracy.

Deepening psychoacoustic modeling is a direction that pursues a higher level of human-centered care. To achieve a more comprehensive assessment of soundscape quality, future research should aim to extend the model's predictive indicators from a single loudness measure to psychoacoustic parameters such as sharpness and roughness, which reflect the "texture" and annoyance of sounds. Building a model that can predict multi-dimensional psychoacoustic indicators would provide unprecedented, refined, and human-centered guidance for urban soundscape design.

To address the lack of temporal analysis, the development of spatiotemporal dynamic models is a crucial direction. By integrating multi-day street view image sequences or video streams and using architectures like Recurrent Neural Networks (RNNs) or Transformers, the model could predict daily and weekly variations in noise patterns, thereby supporting more detailed, time-aware urban planning.

In terms of application, the focus should shift to a model-policy feedback loop. By collaborating with municipal departments, the diagnostic results could be applied in a pilot area (e.g., a specific street segment in Rotterdam) to implement interventions such as increasing green spaces. Acoustic measurements would then be conducted before and after the intervention, using A/B testing to quantify the actual effect of the model-guided policy intervention. This would bridge the gap between theoretical insights and practical outcomes.

To enhance the depth of explanation, causal inference methods could be introduced to separate causal relationships from strong correlations, thereby improving the reliability of policy recommendations. Together, these directions promise to expand the research paradigm into a scalable and adaptable global tool for urban noise management.

7.5. Conclusion

This study does not aim to provide a final tool that can "perfectly predict" urban noise. Instead, it successfully demonstrates a new paradigm for understanding and diagnosing urban sound environments. By translating invisible noise issues into a combination of observable and explainable visual features, this research provides urban planners with a low-cost, high-efficiency "stethoscope." While it may not predict every single loud event, it can reveal the deeper patterns within the urban fabric that lead to long-term noise or tranquility. This advancement represents a solid and inspiring step toward creating a healthier and more livable soundscape, opening up new possibilities for the integration of technological innovation and human-centered governance.

The contributions of this study are multi-dimensional and interdisciplinary, and can be summarized at three levels:

- **Theoretical Contributions:** The most significant contribution is the provision of an effective framework to address the long-standing problem of "micro-macro disconnect" in urban noise management. It was proven that microscopic visual data can be used to infer the micro-scale urban sound environment, which theoretically bridges the gap between macro policies and micro realities.
- **Methodological Contributions:** Two groundbreaking concepts were introduced and validated. The first is "diagnostic mapping," which upgrades the traditional "monitoring" function of noise maps to include the ability to "diagnose" the causes of noise, greatly enhancing the decision-making value of noise assessment. The second is the "dual-metric assessment system," which shifts noise management from a singular focus on physical compliance to a more human-centered concern for soundscape quality by simultaneously evaluating physical acoustic indicators (L_{Aeq}) and psychoacoustic indicators (Zwicker Loudness).
- **Policy Application Contributions:** This study goes beyond theoretical discussion or model

construction. It further proposes a specific, detailed policy translation framework that is closely aligned with real-world regulations. This framework clearly demonstrates how insights from the AI model can be converted into actionable policies under new regulations, such as the *Omgevingswet*, providing urban planners with a powerful and evidence-based decision-making tool.

In summary, this study not only offers a powerful new tool but, more importantly, provides a critical framework for how to responsibly adopt and apply it. This work has paved the way for a future of quieter, healthier, and more intelligently managed cities. This exploration is a beginning, not an end, and the vast prospects it has revealed await further in-depth exploration and practice by future researchers.

Bibliography

Abbott, P. and Nelson, P. (2002). *Converting the UK traffic noise index LA10, 18h to EU noise indices for noise mapping*. Transport Research Laboratory UK.

ADEME and Conseil National du Bruit (2021). Estimation du coût social du bruit en france et analyse de mesures d'évitement simultané du bruit et de la pollution de l'air. Accessed: 2025-05-27.

Atlas Leefomgeving (2025). Kaarten Geluid in je omgeving | Atlas Leefomgeving. https://www.atlasleefomgeving.nl/thema/geluid-in-je-omgeving/kaarten. Accessed: 2025-05-27.

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., and Stansfeld, S. (2014). Auditory and non-auditory effects of noise on health. *The lancet*, 383(9925):1325–1332.

Can, A., Van Renterghem, T., Rademaker, M., Dauwe, S., Thomas, P., De Baets, B., and Botteldooren, D. (2011). Sampling approaches to predict urban street noise levels using fixed and temporary microphones. *Journal of Environmental Monitoring*, 13(10):2710–2719.

Casey, J. A., Morello-Frosch, R., Mennitt, D. J., Fristrup, K., Ogburn, E. L., and James, P. (2017). Race/ethnicity, socioeconomic status, residential segregation, and spatial variation in noise exposure in the contiguous united states. *Environmental health perspectives*, 125(7):077017.

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299.

European Court of Auditors (2025). Special report 02/2025: Urban pollution in the eu –cities have cleaner air but are still too noisy. Accessed: 2025-05-27.

European Environment Agency (2020). *Environmental noise in Europe, 2020*. Publications Office of the European Union.

European Parliament and Council of the European Union (2002). Directive 2002/49/ec relating to the assessment and management of environmental noise.

Evidently AI Editorial Team (2025). Ai regulations: Eu ai act, ai bill of rights, and more. Evidently AI blog. Published January 9, 2025; last updated February 18, 2025.

Gartner, Inc. (2025). Gartner hype cycle model. https://www.gartner.com/en/research/methodologies/gartner-hype-cycle.

Glesser, M. et al. (2021). Mosqito: Modular sound quality integrated toolbox.

Grantham Research Institute (2004). Environmental Management Act 2004. https://climate-laws.org/document/environmental-management-act-2004_c853?q=Environmental+Management+Act. Accessed: 2025-05-28.

Grantham Research Institute (2024). Environment and Planning Act of the Netherlands (Omgevingswet). https://climate-laws.org/document/environment-and-planning-act-of-the-netherlands-omgevingswet_6f48?q=Environment+and+Planning+Act+of+the+Netherlands+%28Omgevingswet%29. Came into force: January 1, 2024; Accessed: 2025-05-28.

Hahad, O., Kröller-Schön, S., Daiber, A., and Münzel, T. (2019). The cardiovascular effects of noise. *Deutsches Ärzteblatt International*, 116(14):245.

Hemmat, W., Hesam, A. M., and Atifnigar, H. (2023). Exploring noise pollution, causes, effects, and mitigation strategies: a review paper. *European Journal of Theoretical and Applied Sciences*, 1(5):995–1005.

Hong, K. Y., Pinheiro, P. O., and Weichenthal, S. (2020). Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data. *Environment International*, 144:106044.

Howarth, A., Pearce, D. W., Ozdemiroglu, E., Seccombe-Hetta, T., Wieringa, K., Streefkerk, C. M., and Hollander, A. E. M. d. (2001). Valuing the benefits of environmental policy: The netherlands. report 481505 024, RIVM (National Institute for Public Health and the Environment) and EFTEC (Economics for the Environment Consultancy Ltd.).

Huang, J., Fei, T., Kang, Y., Li, J., Liu, Z., and Wu, G. (2024). Estimating urban noise along road network from street view imagery. *International Journal of Geographical Information Science*, 38(1):128–155.

ISO532-1:2017 (2017). Acoustics—methods for calculating loudness—part 1: Zwicker method.

Ito, K., Zhu, Y., Abdelrahman, M., Liang, X., Fan, Z., Hou, Y., Zhao, T., Ma, R., Fujiwara, K., Ouyang, J., Quintana, M., and Biljecki, F. (2025). Zensvi: An open-source software for the integrated acquisition, processing and analysis of street view imagery towards scalable urban science. *Computers, Environment and Urban Systems*, 119:102283.

Kephalopoulos, S., Paviotti, M., Anfosso-Lédée, F., Van Maercke, D., Shilton, S., and Jones, N. (2014). Advances in the development of common noise assessment methods in europe: The cnossos-eu framework for strategic environmental noise mapping. *Science of the Total Environment*, 482:400–410.

Kirillov, I. and Bulkin, V. (2015). The mobile system of urban area noise pollution monitoring. In *2015 Second International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T)*, pages 200–203. IEEE.

Klatte, M., Bergström, K., and Lachmann, T. (2013). Does noise affect learning? a short review on noise effects on cognitive performance in children. *Frontiers in Psychology*, 4:578.

Mukim, M. and Roberts, M. (2023). *Thriving: Making cities green, resilient, and inclusive in a changing climate*. World Bank Publications.

Münzel, T., Sørensen, M., Schmidt, F., Schmidt, E., Steven, S., Kröller-Schön, S., and Daiber, A. (2018). The adverse effects of environmental noise exposure on oxidative stress and cardiovascular risk. *Antioxidants & Redox Signaling*, 28(9):873–908.

National Institute for Public Health and the Environment (RIVM) (2023). Road traffic and neighbours identified as main sources of annoyance and sleep disturbance. Accessed: 2025-03-21.

Oosterlee, A. and Zandt, I. (2017). Gezondheidsmonitor volwassenen en ouderen 2016: Belevingsonderzoek naar hinder en slaapverstoring vliegverkeer schiphol. Accessed: 2025-05-27.

Organization, W. H. et al. (2011). *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe.

Penteado, L. D., de Souza, L. C. L., and Christoforo, A. L. (2018). Reverberation time as an indicator for noise maps. *Journal of Urban & Environmental Engineering*, 12(2).

Picard, D. (2021). Torch. manual_seed (3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision. *arXiv preprint arXiv:2109.08203*.

Rijksinstituut voor Volksgezondheid en Milieu (RIVM) and Gemeentelijke Gezondheidsdiensten (GGD)'en (2019). GGD-richtlijn medische milieukunde: omgevingsgeluid en gezondheid - Uitgangspunten en basisadviezen.

Schreurs, E., Verheijen, E., and Jabben, J. (2011). Valuing airport noise in the netherlands: Influence of noise on real estate and land prices. letter report, RIVM.

Smith, M. G., Cordoza, M., and Basner, M. (2022). Environmental noise and effects on sleep: an update to the who systematic review and meta-analysis. *Environmental health perspectives*, 130(7):076001.

Song, L., Liu, D., Kwan, M.-P., Liu, Y., and Zhang, Y. (2024). Machine-based understanding of noise perception in urban environments using mobility-based sensing data. *Computers, Environment and Urban Systems*, 114:102204.

Stibbe (2024). Geluid onder de omgevingswet. Publication / Insight blog. Accessed: 26 June 2025.

Sustainability Directory (2025). Psychoacoustic Metrics. `https://energy.sustainability-directory.com/term/psychoacoustic-metrics/`. Accessed: 2025-05-28.

Verma, D., Jana, A., and Ramamritham, K. (2020). Predicting human perception of the urban environment in a spatiotemporal urban setting using locally acquired street view images and audio clips. *Building and Environment*, 186:107340.

Yaseen, M. (2024). What is yolov8: An in-depth exploration of the internal features of the next-generation object detector.

Zhao, T., Liang, X., Tu, W., Huang, Z., and Biljecki, F. (2023). Sensing urban soundscapes from street view imagery. *Computers, Environment and Urban Systems*, 99:101915.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2020). Places365-CNNs for Scene Classification. `https://github.com/CSAILVision/places365`. GitHub repository; Accessed: 2025-05-29.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2):248–248.

Zwicker, E. and Fastl, H. (2013). *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media.

Acknowledgments

By the time I reached this point in my writing, I had completed my graduation thesis and was about to finish my master's journey. It was a moment worth celebrating, a time when countless emotions welled up in my heart. I had so much to say, and so many people to thank. I shaped these feelings into this acknowledgment—not only as a summary of the past two years, but also as a quiet gift to myself. If my words are not entirely clear—perhaps they are not—then let me tell this story from the very beginning.

I still remember the long preparation and anxious waiting during the application season, and the day I finally chose to study the MSc in Management of Technology at TU Delft. The memory feels as fresh as if it were yesterday. The admission decision came later than promised. In the restless days of early March, I wrote to the NUS secretary to extend my MFE deadline, while also sending more emails to TU Delft, asking about my application. At last, on March 27, 2023, the email arrived, and with it began my story with Delft.

To join the early rounds of housing selection, I had to confirm my payment before April 1. Knowing the delay of international bank transfers, I gathered all the paperwork the very next day, went to the bank with my mother, and completed the transfer. At that moment, my application season finally came to an end, and I could breathe again. In May, when it was time to choose my room, I pored over every review and photograph. Aware of the Netherlands' rainy skies, I wanted as much sunlight as I could get. So the moment my account became active, I chose without hesitation: a panoramic room in Xior, right across from the train station, in the round tower facing due south with windows that could not open. (As I later learned, before moving to my next apartment the following summer, this room's sunlight turned it into a miniature oven—disastrous, but in a way, amusing.) But by then, I was full of hope for the life ahead.

Letian

My flight was set for August 15, from Shanghai to Amsterdam. In the pale light before dawn, I was already dressed, with two large suitcases, a carry-on, and my backpack. My aunt and uncle drove me from Jiading District to the distant Pudong Airport. After more than an hour of check-in and security (including a small incident where a security officer weighed my luggage and found it slightly overweight, prompting me to stuff clothes and electronics into my pockets to meet the standard), I reached the waiting area—and there, for the first time, I met one of the closest friends I would have over the next two years, Letian Li. We had known of each other from a group chat, both headed to TU Delft, and had even chosen seats together weeks earlier.

Letian and I in Rome

It was my first time traveling alone to an unknown country. I felt nervous, even a little lost, but having a friend nearby—even one not yet close—eased that feeling. I read, watched the map on the screen, dozed off, ate airplane meals, and spoke in quiet tones. Fourteen hours later, the Netherlands appeared outside my window. We passed through customs smoothly, went down to the underground train station at Schiphol, and boarded the train to Delft, our six suitcases swaying in the aisle. As expected and unexpectedly, I discovered I had gotten off one stop too early, at Rijswijk, and my apartment was still nine kilometers away. We took the elevator back up to the platform and continued on.

At Delft Station, I said goodbye to Letian, who went toward Delft Campus. I stepped outside at 9:30 pm on a mid-August evening, yet the sky was still bright as afternoon. The cool wind brushed my skin, seagulls crossed the air above me, and as I pushed my luggage forward. Here I was, Delft.

In the two years that followed, whenever a quarter break or even a short gap in the schedule came around (sometimes without any break at all), Letian would suddenly come up with an idea, make a detailed plan, and arrange for us to just set off and have fun! As for that one time we went on a "wilderness survival" cycling trip in southern France. That's a story for another day.

As fellow football enthusiasts, going to stadiums together, watching matches, discussing news, and sharing our rants was always great fun. Thank you so much!

Zhengchu (Tom)

After spending my first night in a completely empty room with nothing but my luggage, the next morning —August 16—I woke up early, already on a European schedule thanks to my last weeks in China (though unfortunately, by now I've shifted to more of a U.S. schedule). Around noon, I walked to the entrance of the X Sports Hall to collect my student card, where I happened to run into Letian and some of his classmates. We strolled around campus for a while, then headed to an open-air bar in the city center for some food and a few beers.

During our chat, the topic of whether there were other Chinese students in our program came up. That's when I searched our WeChat group using "TPM" and "MOT" as keywords and found a classmate named Tom whose profile seemed to match. I sent him a friend request.

When I greeted him, I quickly learned that he was from the previous MOT year—a senior—and, without hesitation, he invited me over for tea that evening. I politely declined at first, but he was so enthusiastic. Although I had some slight worries considering this was the Netherlands, but decided to go eventually. (coincidentally, I would move into the same building in early August this year, as I write this acknowledgment).

Zhengchu and I in Keukenhof (shot by Wu's grilfriend)

Meeting him turned out to be a small adventure in itself. As expected in an unexpected way, I missed my bus stop and had to run a long way back under a bridge, where I met Zhengchu Wu for the first time. Luckily, we really were just drinking tea (kidding). [1] WuGe turned out to be a very normal yet interesting person. We talked about everything—from his past life experiences to advice on my studies, to the local lifestyle, leisure spots, and small life hacks. Another MOT senior, Yuncong Liu, also lived there and was due to return to the Netherlands soon. Together, the three of us—the future "MOT trio"—later treated me to dinner.

When WuGe heard I was going to The Hague the next day to accompany some new friends for their residence permits, he mentioned that he knew a great spot nearby. In his exact words: "That place is awesome. I've been there twice recently, and I still want to go again." The next afternoon, we met him at the station near the location on the map, and he led us to the beach in The Hague. The weather wasn't too hot, the sun was out, and the bar by the sand made for a perfect spot to relax. In the evening, he took us for Cantonese food.

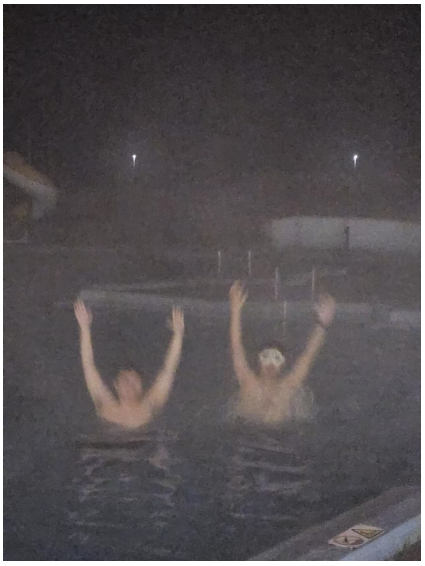
And that was the beginning of our MOT group's habit of going out and enjoying life in the Netherlands. Even back then, I knew I was starting to like my days here.

What came after were countless memories: WuGe became my coach and made me a workout plan, we traveled to Iceland, celebrated New Year's Eve, went to the Olympics, drove to Walibi, and so much more. In both life and academics, his help and support meant a great deal to me. This May, he chose to return to China for work. I am deeply grateful that I met WuGe here in the Netherlands.

Yuncong (Spike)

On August 22, WuGe invited me to a restaurant in Delft city center. It was part of a plan he and another friend had booked in advance, and there I finally met the legendary Yuncong Liu (Spike), known for "feeding" WuGe all year with his excellent cooking. I'll refer to him here as CongGe.

When I first met CongGe, he had a slight sick, a broad smile, and a solid build. Coincidentally, his university city and my hometown are both in the same province—Shaanxi. We hit it off right away. In September, when WuGe left us to visit his girlfriend in the UK, I went to CongGe's place for the first time for hotpot. I bought quite a lot of ingredients, he prepared plenty as well, and together we finished everything! From then on, I kept to CongGe's rule—whenever I had the time, I could go to his place to cook and eat.



Yuncong and I in Iceland

[1] For convenience, I'll refer to him from here on as "WuGe"—in Chinese, "Ge" roughly means "bro/man," a friendly way to address someone by combining part of their surname or given name with "Ge."

Over the next two years, I lost count of how many times I enjoyed his amazing meals. I felt truly happy. We also went on many outings together—sometimes just heading to a restaurant we suddenly craved, other times hiking in the Netherlands, enjoying the journey, the exercise, and our shared interest in outdoor gear and electronics. In the second half of my second year, when WuGe stopped showing up at the gym, CongGe and I began training seriously together, and I gained a lot from it (though in the end, he returned to China for good).

In these two years, I witnessed CongGe's ups and downs—both physically and otherwise. Our trips together were always spontaneous: once we decided on a destination, we went. We traveled west to Munich for a Coldplay concert, taking an overnight Deutsche Bahn train; we took a cruise during our Nordic trip and trekked through muddy trails in Bergen. I'll never forget how, during almost every trip, CongGe would pull out his earphones to check in or video call with his partner back in China, something that left a vivid impression on me.

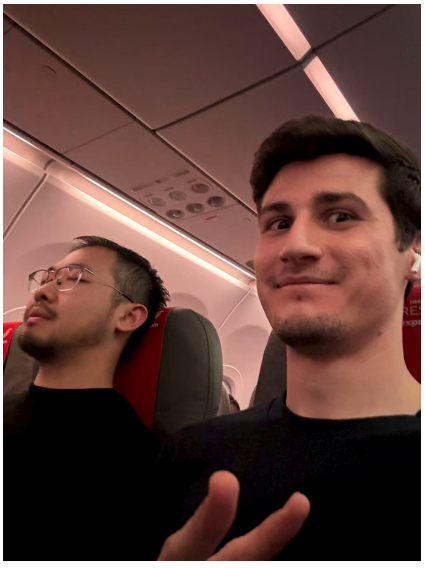
Life with CongGe was always fun and full of laughter, and his advice was truly valuable to me. I'm grateful to him and to the "legacy" he left me here in the Netherlands. I wish him all the best in his work and life in Xi'an.

Isa

At the end of August, on August 30, 2023, our MOT program held a welcome event for new students. Like everyone else, I went to the TPM building to take part. Earlier that day, I had gone to the city center cinema to watch *Oppenheimer*, my first English-language movie in a theater. The subtitles were neither in Chinese nor in English, which felt novel, although it did make the film harder to follow. As soon as the movie ended, I got on my bike, followed the navigation, and headed straight to the TPM building.

Walking through the entrance, I found my tall, bald mentor Eric. After a brief chat, he gathered the other members of our group so we could introduce ourselves. When we heard there would be free coffee and snacks, Isa Ouz said, "Who else wants some free stuff?" I immediately ran over to pick something out, joking, "As long as it's free, I love it no matter what it is." That was how we started chatting happily together.

In the two years that followed, we shared many experiences: attending classes, working together on group projects, teaming up again for FAIP where I handled coding while he presented, traveling to Madrid's UPM for both leisure and coursework exchanges, and helping each other move. I am grateful for Isa's help, for the times we had deep conversations, and for wandering the streets of Madrid talking about life. His support over these years meant a lot to me.



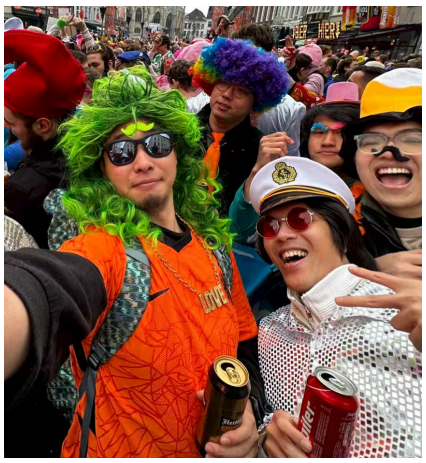
I and Isa on the flight to Madrid

More than once, Isa told me, "Your English improved a lot!" Hearing that made me realize how much my casual conversation and joking skills had grown. Isa is both a natural leader and an outstanding writer, someone who naturally earns people's trust. I look forward to seeing him accomplish great things in the future.

Khanh

On that same day, during the same event, I noticed an East Asian-looking face and a man who happened to share my surname. I tried speaking Chinese to confirm my guess, but I was completely wrong. This was Khanh Ma, who introduced himself as being from Vietnam. It turned out that "Ma" is also quite an uncommon surname in his country, which made the coincidence even more interesting.

Before the semester started, we went for a walk around the Delft lake area, chatting about our academic backgrounds, past life experiences, our views on relations between our two countries, and even some politically sensitive topics. He was a steady, thoughtful, and interesting person.



I and Khanh in Breda Carnival

We continued teaming up in some courses, working together with great ease. We also went to Carnival together and enjoyed casual drinks in Leiden. I am very grateful for Khanh's help. As one of the few "true brothers" from a neighboring country to my own, I hope that in the future he will be the reason I have influential connections in Vietnam—haha.

Mihkel

The blond, blue-eyed man also appeared on that same day. He wasn't in my group, but during the beer session at the end, while I was in the café area picking up my bag and chatting, he happened to be standing to my right. Naturally, we started talking. It was clear that this Estonian man, Mihkel, carried a certain coolness—not overly warm, but undeniably beautiful.

It wasn't until Q2, when we were paired for the Research Methods course, that we began working more closely together. Since our academic backgrounds were similar, we decided to team up. As our conversations grew, I discovered his endearing and humorous side. Perhaps, to some extent (I'm not sure how much), I influenced his decision to apply for a half-year exchange in China during his second year.

By coincidence, in December 2024, after a few days of hiking on my first trip back to China, I stopped by Beijing and met him on the Tsinghua University campus. It was quite a fun twist of events. Now his Chinese girlfriend has also come to the Netherlands, making the whole story feel even more like fate—an amusing and memorable connection.

Mihkel and I in TPM

Fahmi

I met Fahmi in class that September. He isn't very tall, and at first glance I thought he might be a prince from a Southeast Asian country. He is a devout believer, yet also incredibly endearing, often making jokes about himself. Every time we met, he would greet me warmly, always with a playful remark.

At the end of each quarter, he was always the first to ask about my academic results, trying to "compete" with me over grades. It wasn't until the Indonesian event this February that I discovered he could dance! Over these years, I've been grateful for the joy and laughter Fahmi has brought into my life.

Fahmi and I in inDelftnesia

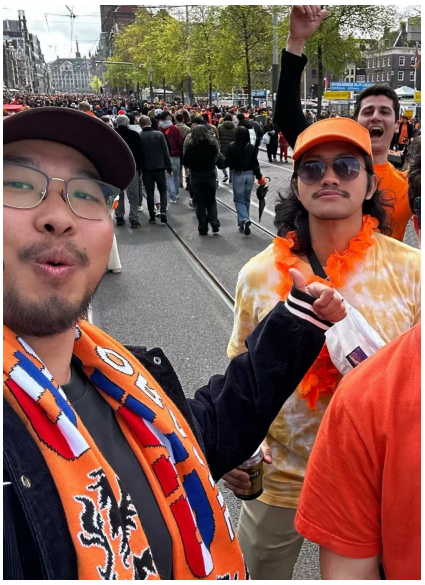
Bayu

Mafia—that's the only way I can refer to Bayu. It was my very first impression of him, and it has always matched his image and aura: an Indonesian mafia boss. Muay Thai, muscles, cool tattoos—I often teased him about it.

We really became closer during the Mid-Autumn Festival in 2023, when I invited him and others to have Chinese food together and shared the premium mooncakes I had spent the afternoon hunting down in

The Hague. He was also my study partner. When I failed the first course of my life (Leadership) and learned that he also had to take the resit, we studied side by side late into the night at Echo, Aula, and the library. In the end, we both achieved our goals.

Mafia invited me many times to his place to hang out or have a meal, but unfortunately, our schedules rarely matched—my late-night routine and his time training friends in martial arts didn't align—so I never made it, something I regret. Still, we managed to join plenty of other activities together. I also can't forget one particular evening: late at night in Indonesian time, I was at home in Delft when I suddenly got a call from him begging me to help fix his code. That's when I learned he was actually on vacation in Bali, but couldn't get it to work and was on the verge of failing.



I and Bayu on King's Day in Amsterdam

Thank you, Mafia, for two years of protection and friendship. And I should add—beneath it all, he's truly kind and friendly.

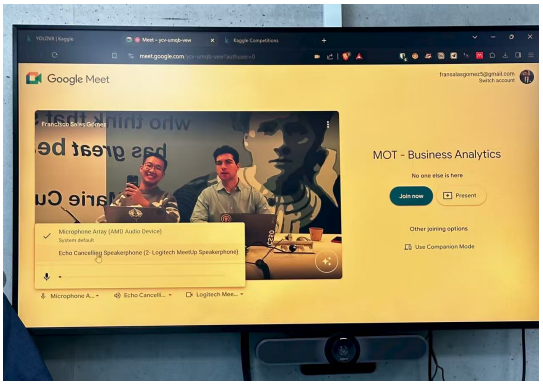
Fran

On September 20, 2023, during TPM's Freshers' Week, the faculty took us to explore the Delft lake area, visit museums in The Hague and Rotterdam, and enjoy a large sightseeing cruise. I remember at The Hague's history museum, Mafia arrived late together with Francisco. He was strikingly handsome, with chiseled features, deep green eyes, and the presence of an ancient sculpture—wearing a suit yet carrying an air of ease. It wasn't until we were on the deck of the Rotterdam boat that I greeted him for the first time and we got to know each other.

Later, in class, we exchanged film recommendations and ended up watching *God's Crooked Lines*. I also invited him to our Mid-Autumn Festival dinner with Mafia and Naffi. I grew very fond of this man with his dry humor. As fellow football fans, and with Spain as my top European national team, Fran and I always had a great time watching matches together. For the Euro final, we stood among a sea of Spanish supporters at X, celebrating together, and that day gave me several of my favorite social media profile photos in a Spain jersey. (The only "flaw" was that, being from Madrid, he was—unsurprisingly—a Real Madrid supporter.)

A funny little episode happened in late 2024 during El Clásico. Fran invited me to watch the match at X with his girlfriend. At halftime, we moved to my place to watch the rest. On the way, he expressed his confidence and excitement, but after Barcelona scored four goals, his changing expressions were so dramatic that I could barely hold in my laughter. After the match, I joined his girlfriend in consoling him.

I've been lucky to work with Fran on coursework—he is steady, reliable, and an excellent coordinator.



I and Fran as interviewers on Valentine's Day 2024

I'm grateful to him, our guide to the European Space Agency, the runner and the CEO.

Many More MOT fellows

There are so many more friends I met through the MOT program that I want to thank. There is Oscar, my stylish Black Brazilian-Dutch buddy who always tried to help in group work but somehow always ran into problems (and who has one photo where his side profile looks strikingly like Chris Paul). There is Ketill, the Icelandic tough guy with a taste for sharp American slang. There is Anuj, small in build yet quick with jokes, whose birthday let me witness a scene straight out of a movie—dozens of people dancing to Indian songs. There is Bhavesh, precise and reliable, a pleasure to work with, and practically the poster man for X.

I also want to thank Felix, my elegant yet athletic Dutch friend who has been to Tibet, a place I have not yet reached myself. Then there is Stefan, our ultimate tough-guy dad and pilot. I remember Peng Lee, my fellow Taiwanese friend, who once took me bouldering, and later invited me and Isa, along with our FAIP teammate Yfka, to dinner. And of course, Hijme and Luis: the tall, cheerful Dutchman and our dependable, skilled Peruvian finance expert.

Shoumeng, a junior in TPM studying in CoSem, I'm grateful for our shared experiences watching matches and exchanging our complaints. Thanks to her, I was able to make my pilgrimage to Westfalen. To this day, my unbeaten record when supporting a team in person and her winless record when she does so remain a legendary part of our football-watching history.

They are all such wonderful people, and I feel deeply grateful to have known them.

Chen, Haoyu & Lai

Chen Bao and Haoyu Wang (BC and WangShen [2]) were friends who joined us in the early and middle stages of our fitness journey, and they became important companions in my life in Delft. BC is an excellent cook, and back in our first year when we lived in the same apartment building, I often got to enjoy his dishes. He would sometimes come over to my place to watch the NBA, and after workouts, we would have protein shakes together. He, in WuGe's social group for gaming, remains a mysterious person and never shows up in reality.

WangShen, who was in the same cohort as WuGe and CongGe, later became my "apprentice" in the gym. After graduating, he moved out of Delft, but we still met up from time to time for single-day trekking trips in the Netherlands, making life both active and enjoyable.

Lai Wei, once WuGe's great benefactor in the early days and his former top gym apprentice, is now pursuing a PhD in EE and has also become a good friend of ours. I am truly grateful to them.

[2] Shen in Chinese means brilliant person.

Lai, Zhengchu, Chen, and I in front of the Schiphol Airport the day Zhengchu left us forever for China

Xiaolei

Xiaolei, a beautiful and fascinating woman. On October 12, 2023, the day of the Xior apartment community event, I went to the ground floor common room right on time. There was plenty of beer and snacks, and I happily chatted with neighbors I hadn't met before. Around eight or nine in the evening, while I was talking with a new Indian neighbor in the lobby, a Chinese girl walked up to ask if I could take a photo of her and her friends. I naturally agreed, and afterward we started chatting. She told me about the drawing workload at the BK faculty, and I found her cheerful and humorous.

In my second year, after I moved to a new place, WangShen sold me some second-hand items that were still in the neighboring apartment building. When I tried to find out who lived on that floor, I discovered it was her. She came to ask if the table in the common room was mine and helped open the door. Together with WuGe and CongGe, I carried the items back to my place. Later, by chance, she began going to the gym with me and became my top female apprentice. I trained her, shared protein shakes, and played the newly released *Remnant II* together.

We joined Khanh, CongGe, Mafia, and other friends for Carnival in Breda, had the first "Ma family" gathering (yes, she also shares the Ma surname, though her hometown is thousands of kilometers from mine), went on a one-day hike with WuGe, and cooked and dined together at my place with friends. Now she is my girlfriend, and as I write this acknowledgment, we have just finished moving into our new home.

I and Xiaolei at the Imagine Dragons concert in Amsterdam

I am grateful for Xiaolei's presence, which gave me emotional support in this most anxious year. I am thankful for this relationship, this chance encounter, and the plans and hopes we now share for the future. A special thanks to Xiaolei—thank you.

Heyuan

On January 7, 2024, I had almost forgotten that Christmas break was ending that weekend when we arranged to go to Haarlem for a simple trekking trip. That was when I finally met my "good son," Heyuan Huang—by then perhaps already "Mss" (if such a term existed). We had known each other for many years, always calling each other "father" and "son." She had come to the Netherlands two years earlier than I to study for her master's at Erasmus and had been working for quite a while. Since arriving, I had always been too busy to meet her (vice versa), but when we finally did, it was exactly as expected—still my good son.

She has been a great help to me, giving me advice about life, sharing analysis and witty complaints about various matters, and understanding the situation in China in ways that closely match my own perspective. She even helped me buy football jerseys, and after she got a car when moving to a new place, she often planned fun activities for us, including but not limited to shooting practice and climbing parks.

I and Heyuan at the Fun Forest in Rotterdam

I am truly grateful to Heyuan for her help and support since I came to the Netherlands. Thank you, my good SON!

USTCers

To my dear friends from the University of Science and Technology of China (USTC), I am truly, deeply grateful to you all. Coming alone to a foreign country, it was your familiar companionship and conversations that helped me through so many days.

KenYe, who taught me the basics back in our undergraduate years, and later, during my first half-year in Delft, video-called me many times. Even after working until 9 or 10 p.m. every night at Huawei, you still listened to my complaints—thank you for that.

Yuhao and Binghan (BingBing) welcomed me when I landed in Shanghai, picking me up and taking me out to eat. Later, Yuhao and I continued our late-night suppers and board games in Hefei, while BingBing, after finding a new job, went with me to a temple to pray for blessings before I returned to the Netherlands.

Hongwei, our travels together were just as spontaneous as they were in our undergraduate days, full of stories and new experiences. Yuxuan, my incredibly handsome former roommate, was one of the first people I met in Hefei after arriving (though in my excitement, I accidentally broke the custom mug I had brought as a gift). Hailin (LinJie [3]), my other "good son," welcomed me during my trip back to China. After resting in Shanghai, I immediately traveled to join her and her boyfriend Xiao Hong for a three-day hike in the Taihang Mountains. I miss her endlessly cheerful, infectious laughter. She is heading to Hong Kong for her PhD and has a bright future ahead.

[3] Similar to Ge use before, Jie is for famale.

I and Hailin in South Taihang Mountain

Unfortunately, I didn't get to see Changqing Zhu (ZhuGe), who has been deeply focused on his studies in the south. But I did reunite with my unbeatable brother, Ziang Wang (WangGe). In early April this year, when I received the news that our mutual best friend, Mardandan (Captain Mai), was getting married in Xinjiang, I checked flights immediately and booked my ticket within an hour. I first flew to Shanghai to see WangGe, who then took leave from work so we could fly six hours together to Xinjiang. At our brother's wedding, I shed tears of happiness. The other guests, upon learning I had traveled 25,000 kilometers to be there, treated me as an honored guest. Between 40 shots of strong baijiu and my attempt at Xinjiang dancing, it truly felt like a celebration. Seeing WangGe again and talking with him brought back so many memories and emotions.

I, Mardandan with his wife and Ziang at the wedding

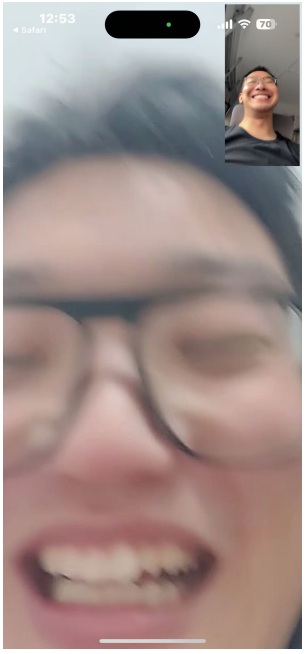
I am very thankful to them all.

Older Friends

To my old friends, whom I have known for around a decade or more, I am truly thankful.

Tianhui, my old deskmate, often chatted nonsense with me during my master's years, making me miss the mischievous "devil's whispers" we used to exchange in class.

ShiGe (whose former name really was "Shi Ge," so he's different from the others mentioned earlier) often debated with me about the historical rankings of football players. We had a great time together at Universal Studios Beijing. Unfortunately, for various reasons, he found a better path, so it seems I won't be seeing him in Europe for a PhD at EPFL.



Tianhui and I on Valentine's Day

ShiGe and I in Beijing

A special thanks to them both.

Lion & Sander & Sander

In September 2024, during the FAIP course, Isa and I chose topics that matched our interests and were assigned to the noise group. That was when we first met Lion, our mentor and a PhD student. In that project, after several twists and turns, we finalized our plan to model, cluster, and predict magnitudes using the New York noise complaint dataset. Lion helped us a great deal, and through this work I developed a strong interest in noise research.

After our final presentation, Lion asked whether we would be interested in doing our thesis in his lab, with several possible topics available. In December of the same year, I decided to continue working with the stylish Lion on a noise-related project—this became the foundation of my thesis.

Lion and I at the last meeting in his office

We met many times throughout the process. Lion gave me valuable advice, guided brainstorming sessions, and provided practical direction. He even helped me purchase an Insta360 camera for data collection, which I admittedly also used over the following months to record my travels. I am deeply thankful to Lion for his patience, thoroughness, and significant contribution to my graduation thesis.

Sander and Sander, both professors, were my supervisors. The first, Sander van Cranenburgh, as my chair supervisor, gave me insightful feedback and inspiration during topic selection and lab meetings. I greatly appreciate the opportunities he provided. The second, Sander Smit, a familiar face in the MOT program who had taught us several courses (I once even applied to be a teaching assistant for his Business Analytics class, though unsuccessfully), was always humorous and gave me well-targeted advice.

Family

I am grateful to my family for their support and for making it possible for me to study in the Netherlands. Thank you so much.

Me

I have so much I wish to say to myself, so many fleeting moments I long to recall, so many regrets that can never be undone, and so many choices for which I owe my gratitude to the person I once was. I tried to borrow a few words to express the feelings within me:

> "*Life may not be as good as you imagine, but neither will it be as bad as you fear. Human fragility and resilience both exceed our own imagination: sometimes a single fragile word can bring tears streaming down your face; sometimes you discover that, gritting your teeth, you have already walked a very long road.*"
>
> "*He does not believe that the long night is coming, for the torch is already in his hand.*" —Guan Weijia, commentary on Kevin Durant's game-tying shot at the end of regulation, 2021 NBA Eastern Conference Semifinals Game 7, Brooklyn Nets vs. Milwaukee Bucks
>
> "*In truth, a real farewell has neither the long pavilions nor the ancient roads, nor the urging to drink one last cup of wine. It is simply that, after such a midday, certain people and certain things are left behind in memory. The day of parting will surely come, and what you miss is not the picturesque scenery of that time, but every person who shared those scenes with you. What you cannot forget is not the days themselves, but the passion you held in your youth.*" —Commentator Yu Jia

All in all, thank you, Ma Yuxiao. As always, I will end my master's degree with the same motto as before:

> *In a world steeped in absurdity, stand as the maverick of my own story.* 在荒诞的世界里做自己的孤胆英雄。

I, Yuxiao

A

Supplementary Model Outputs and Visualizations

This appendix shows the result figures in the main chapters.

A.1. GUI Interface

With the help of Github Copilt, I designed and transfered all functions into the user-friendly GUI interface (See Figure A.1, A.2).

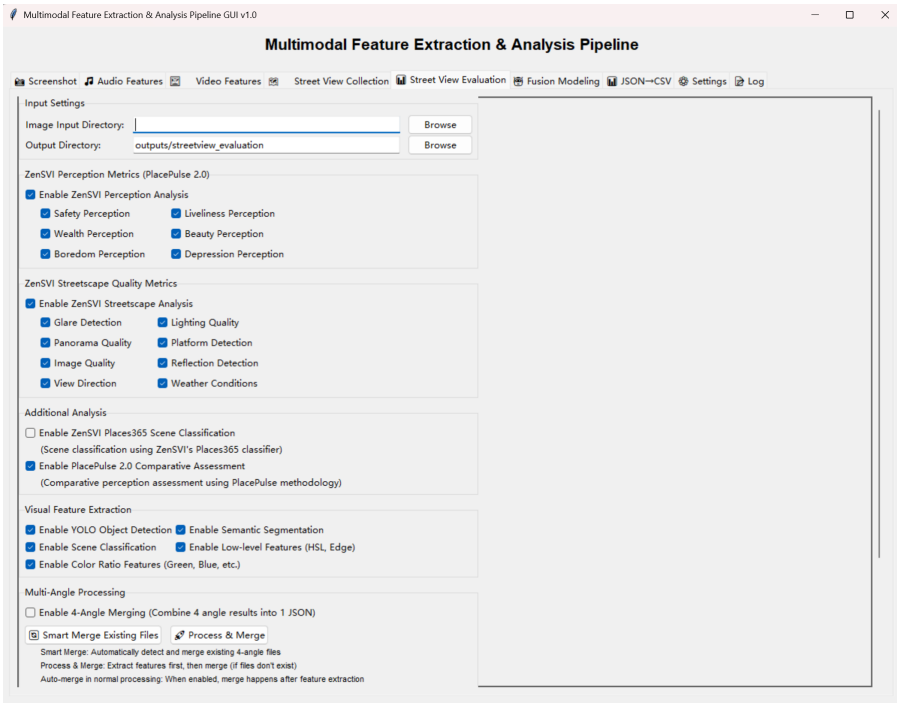


Figure A.1: GUI Interface for Evaluation

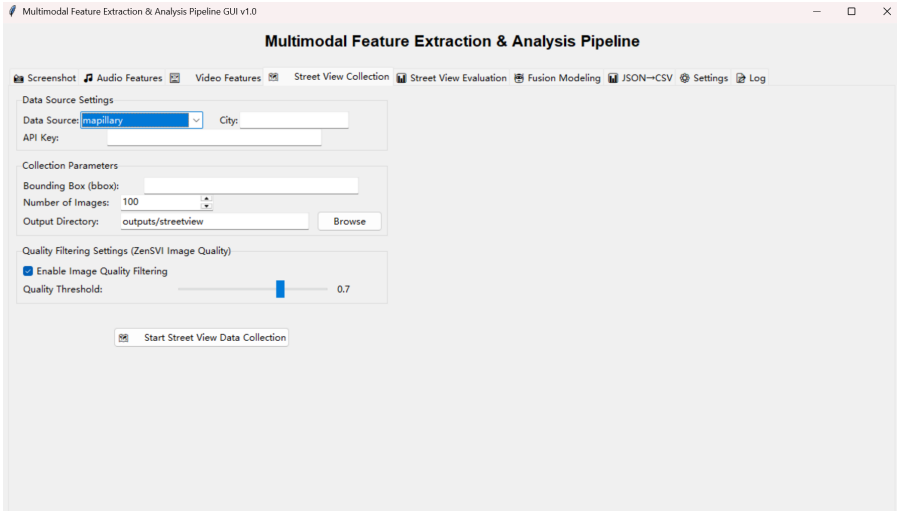


Figure A.2: GUI Interface for Getting StreetView Images

A.2. Exapmle of SVI in the Hague

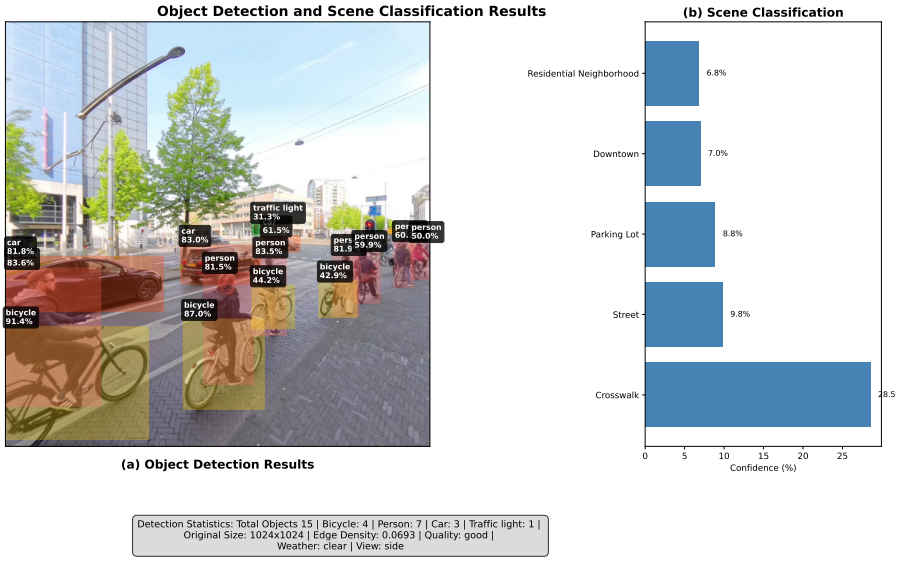


Figure A.3: Detection & Classification from the Hague Street View

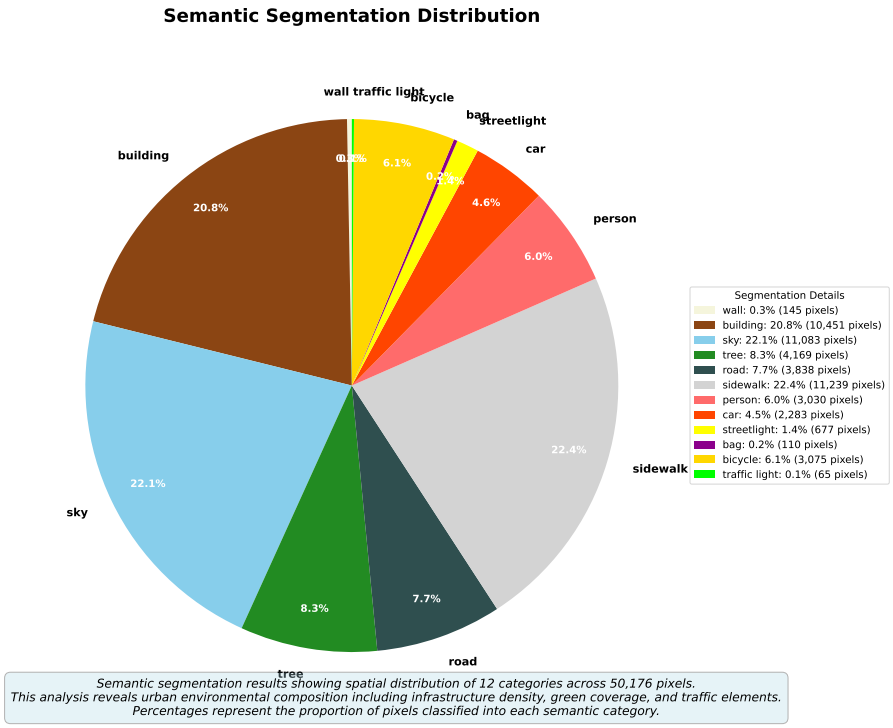


Figure A.4: Segmentation from the Hague Street View

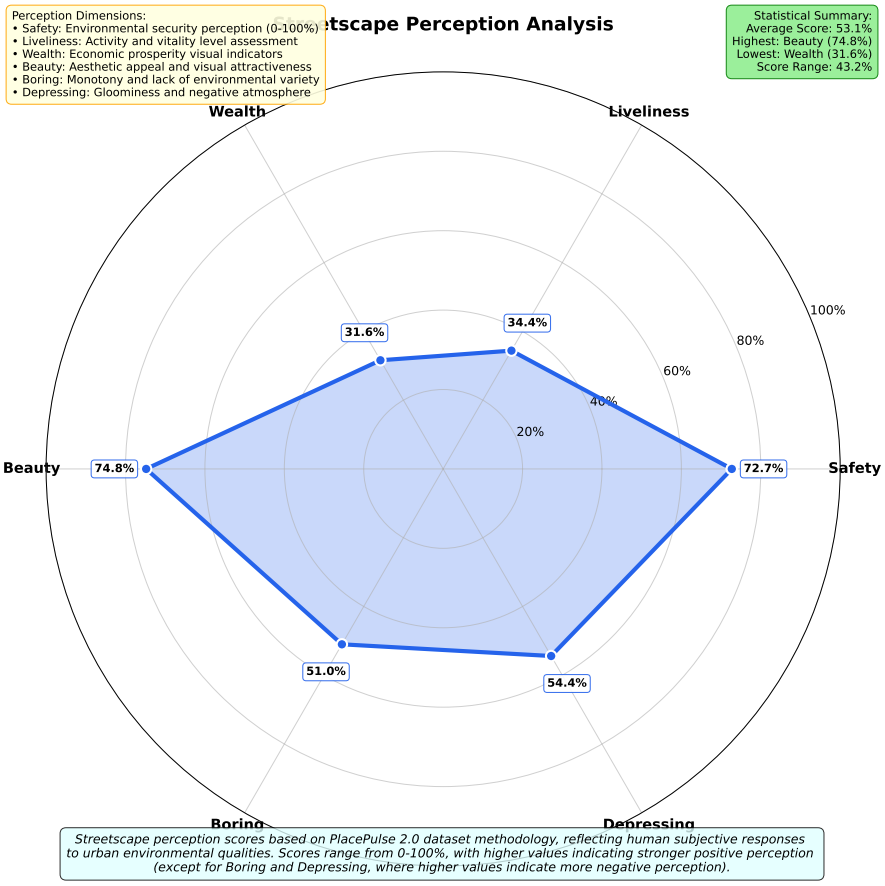


Figure A.5: Streetscape Radar Chart from the Hague Street View

A.3. PCA and Feature

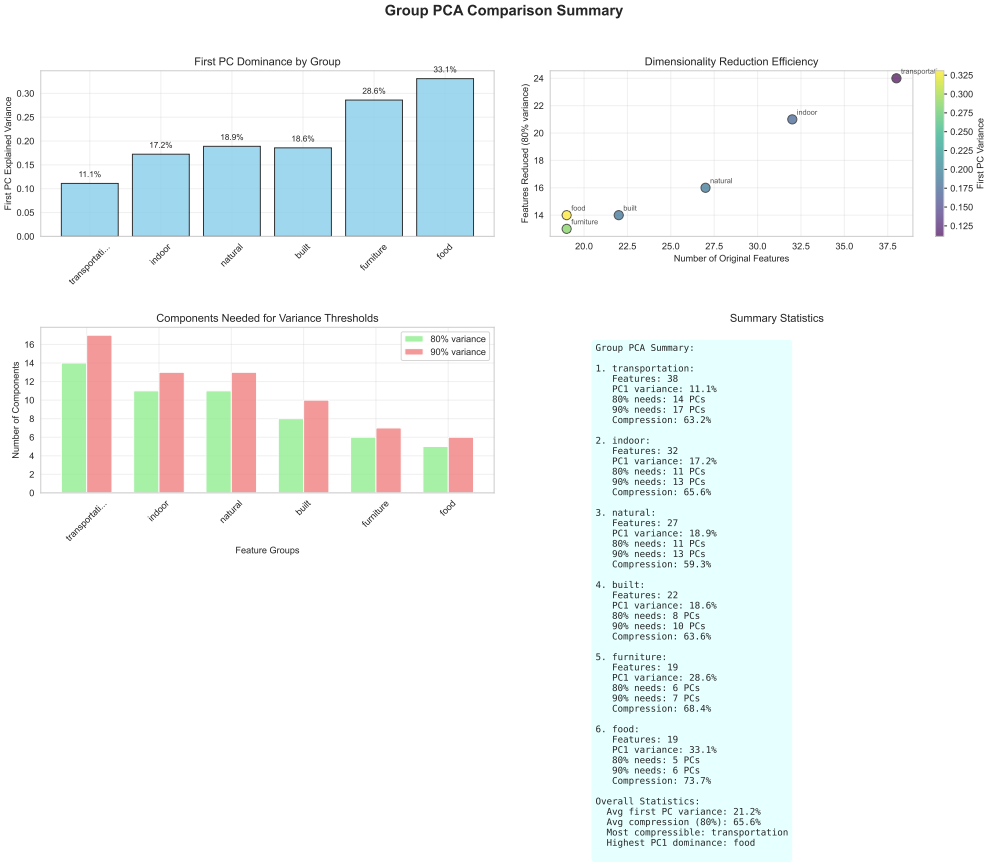


Figure A.6: PCA and Group Analysis

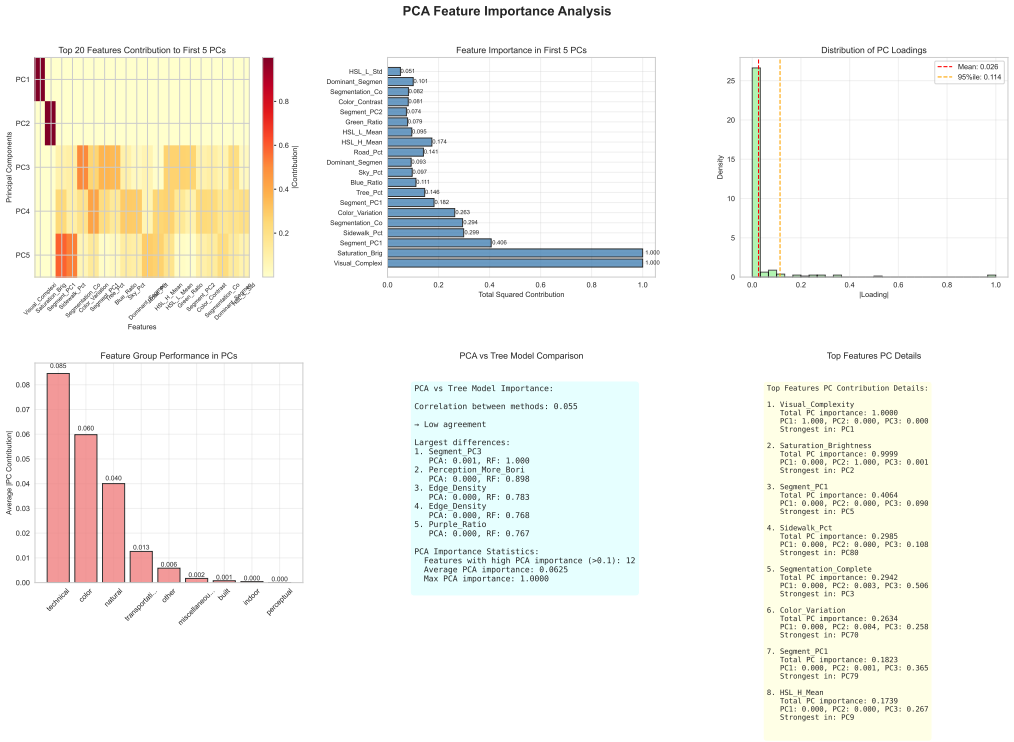


Figure A.7: PCA and Group Analysis 2

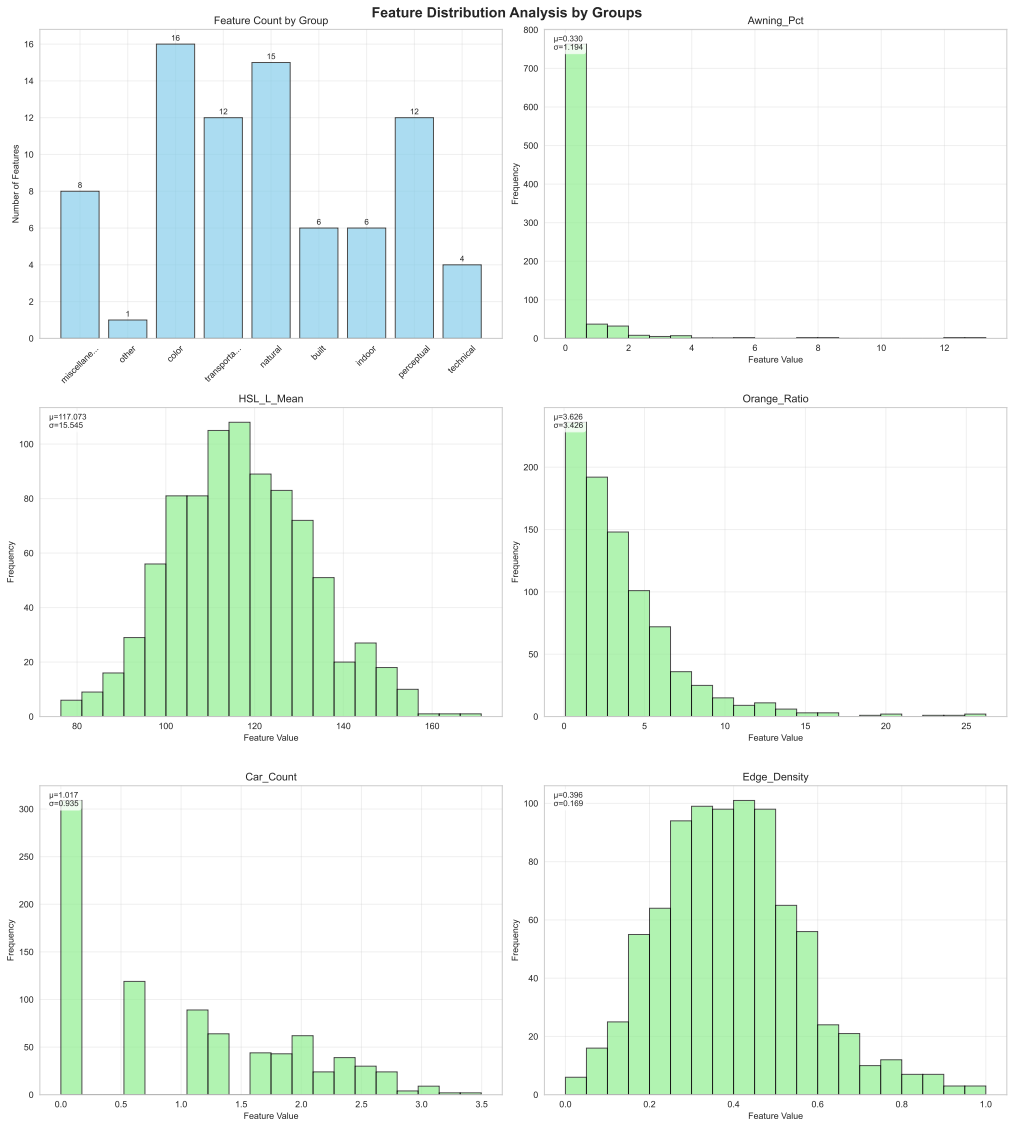


Figure A.8: Comprehensive Feature Analysis 1

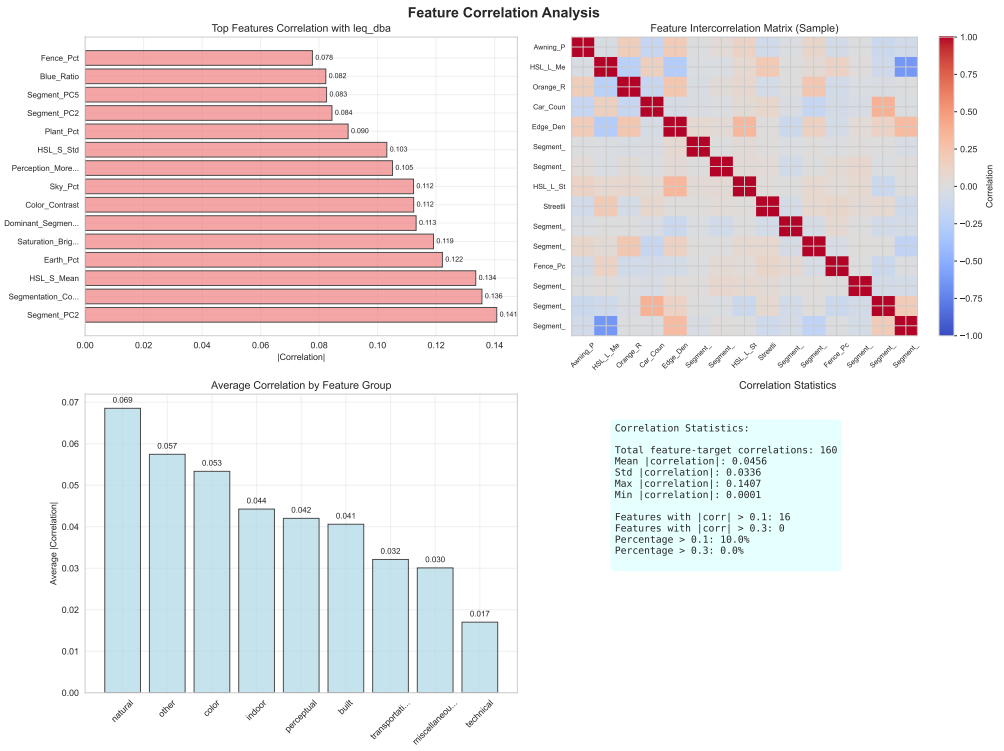


Figure A.9: Comprehensive Feature Analysis 2

Group	Count	Mean	Std	Var
color	16	535.268	197.886	573822.178
natural	15	6.475	6.931	79.501
transportati...	12	6.848	4.805	35.975
perceptual	12	0.366	0.068	0.005
miscellaneou...	8	3.811	2.493	20.173
built	6	0.018	1.027	2.255
indoor	6	0.057	0.273	0.120
technical	4	87606.320	395568.523	625896970186.870

Figure A.10: Comprehensive Feature Analysis 3

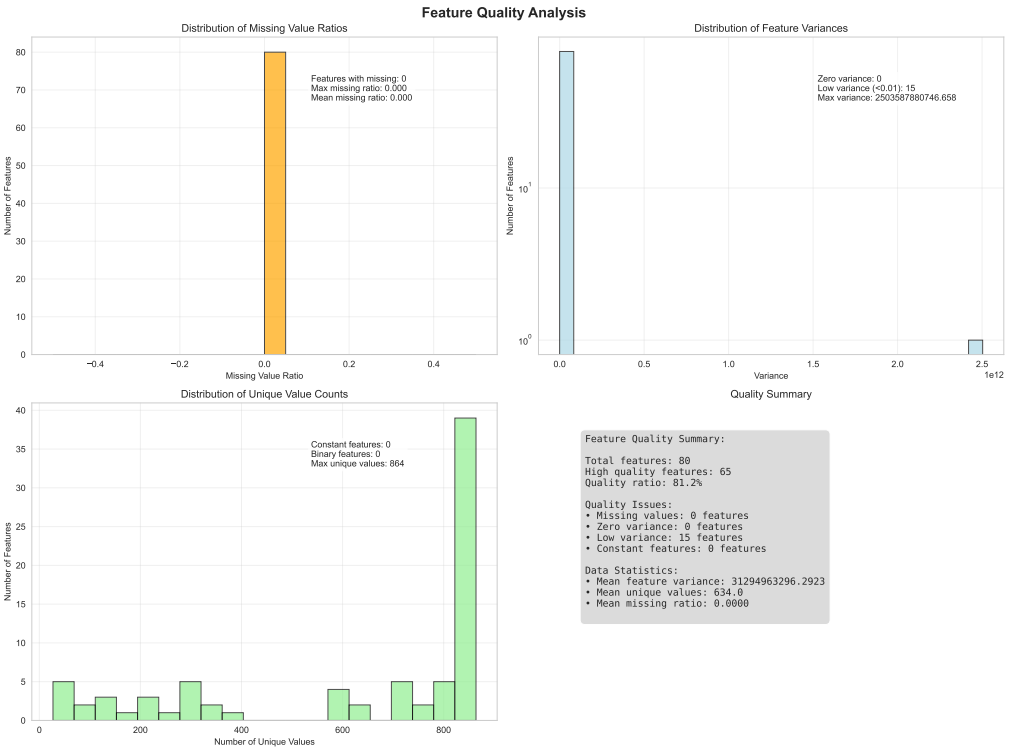


Figure A.11: Comprehensive Feature Analysis 4

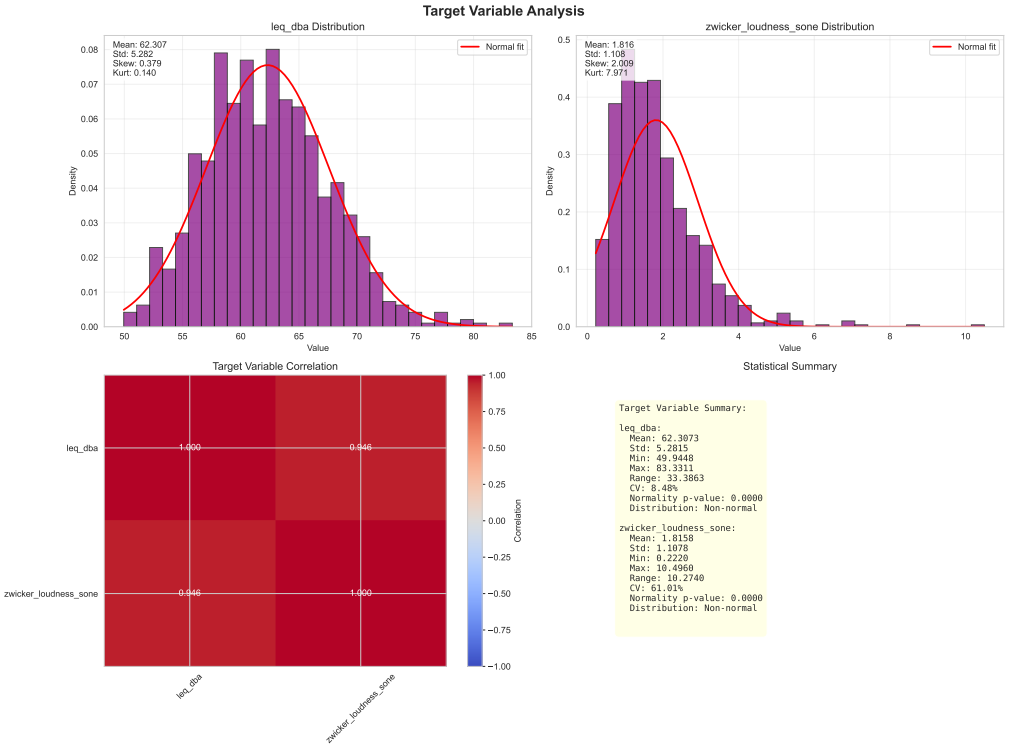


Figure A.12: Comprehensive Feature Analysis 5

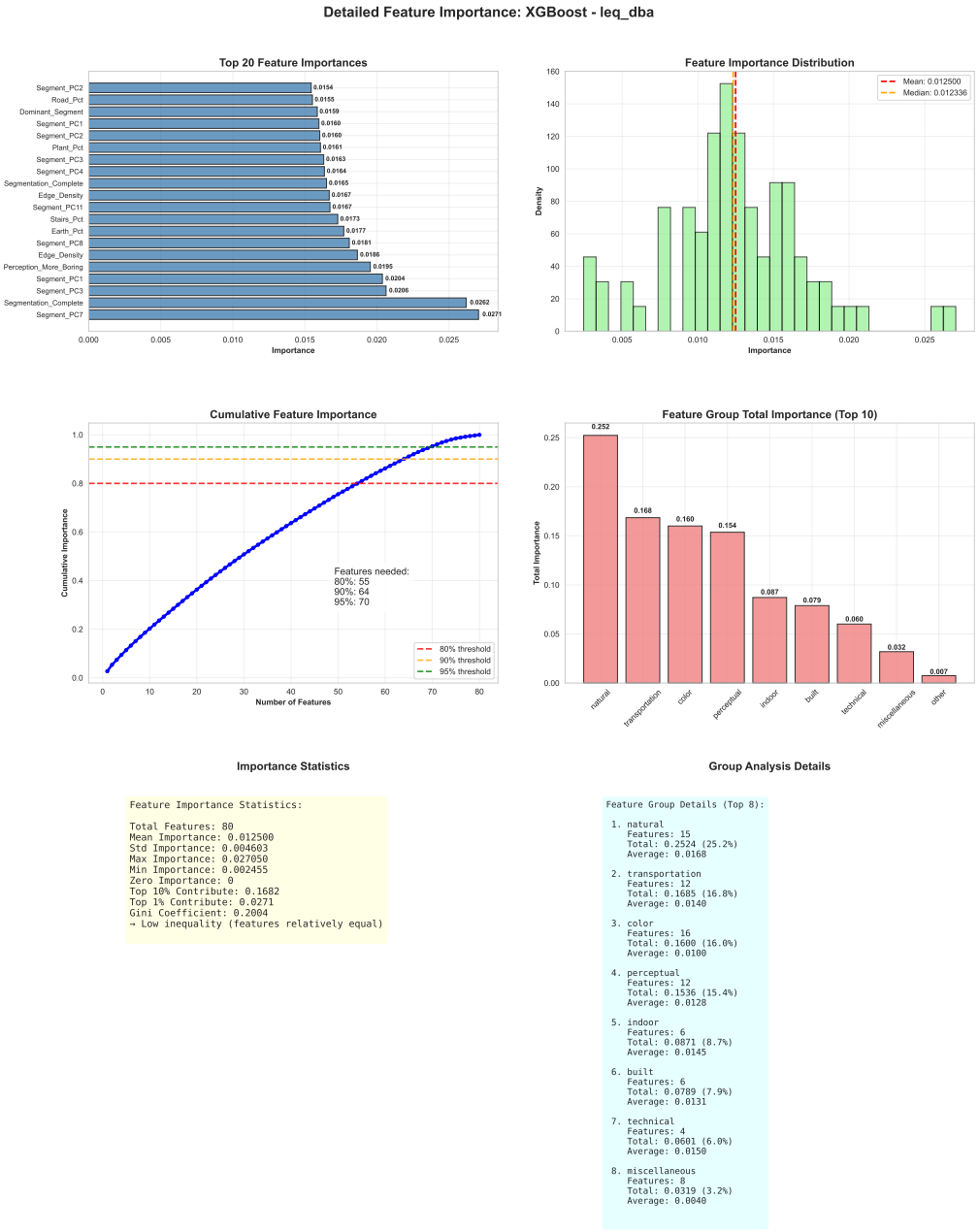
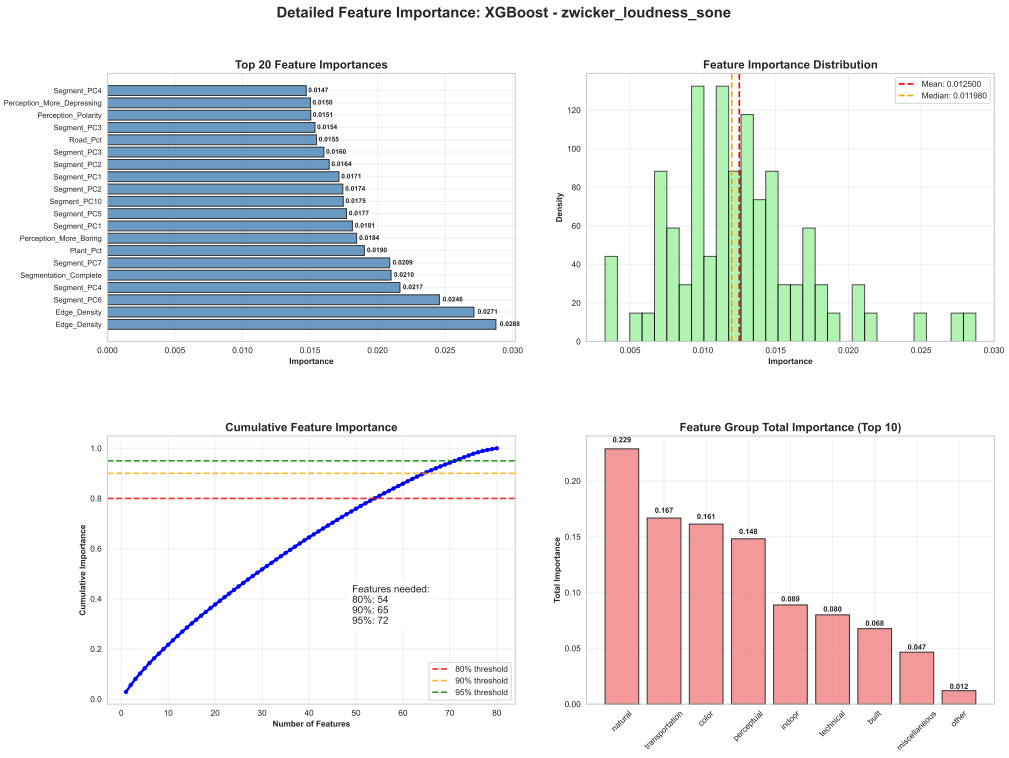


Figure A.13: Comprehensive Feature Importance XGBoost



Importance Statistics

```
Feature Importance Statistics:

Total Features: 80
Mean Importance: 0.012500
Std Importance: 0.004851
Max Importance: 0.028756
Min Importance: 0.003271
Zero Importance: 0
Top 10% Contribute: 0.1815
Top 1% Contribute: 0.0288
Gini Coefficient: 0.2103
→ Low inequality (features relatively equal)
```

Group Analysis Details

```
Feature Group Details (Top 8):

1. natural
   Features: 15
   Total: 0.2287 (22.9%)
   Average: 0.0152

2. transportation
   Features: 12
   Total: 0.1667 (16.7%)
   Average: 0.0139

3. color
   Features: 16
   Total: 0.1613 (16.1%)
   Average: 0.0101

4. perceptual
   Features: 12
   Total: 0.1481 (14.8%)
   Average: 0.0123

5. indoor
   Features: 6
   Total: 0.0889 (8.9%)
   Average: 0.0148

6. technical
   Features: 4
   Total: 0.0800 (8.0%)
   Average: 0.0200

7. built
   Features: 6
   Total: 0.0676 (6.8%)
   Average: 0.0113

8. miscellaneous
   Features: 8
   Total: 0.0466 (4.7%)
   Average: 0.0058
```

Figure A.14: Comprehensive Feature Importance 5

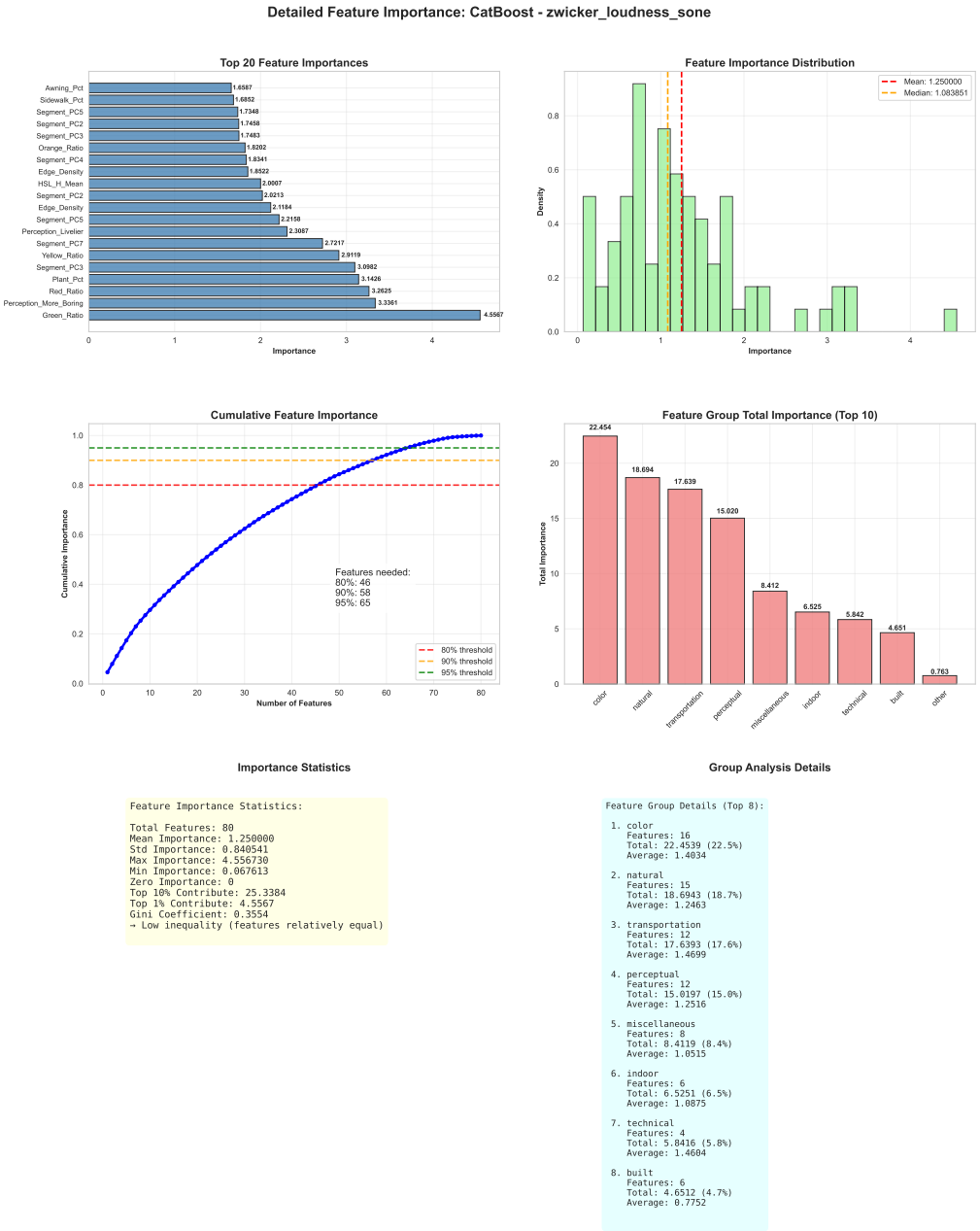


Figure A.15: Comprehensive Feature Importance 6

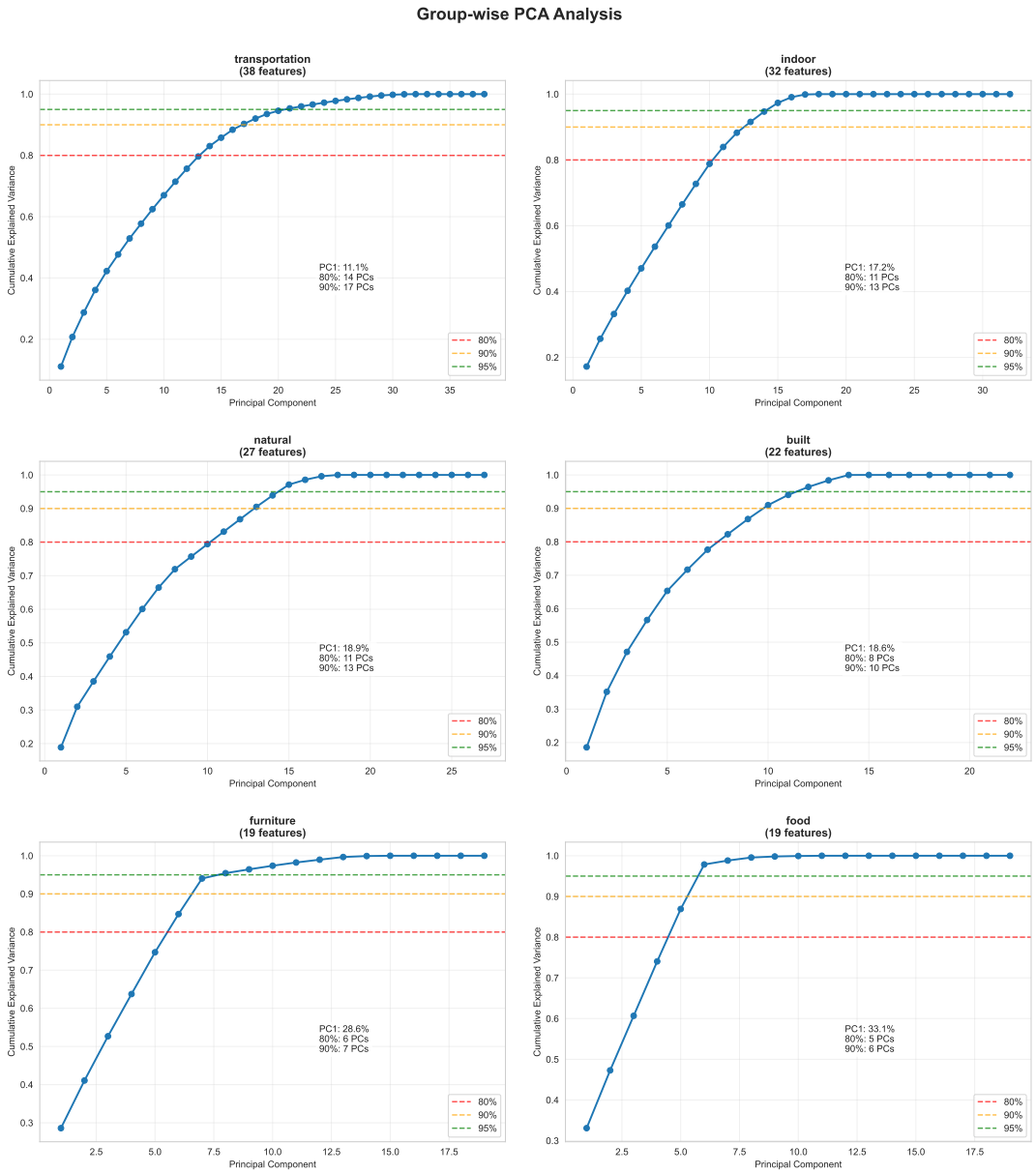


Figure A.16: Grouped PCA Analysis Results - Cumulative Explained Variance Curves for Selected Feature Groups.

A.4. SHAP

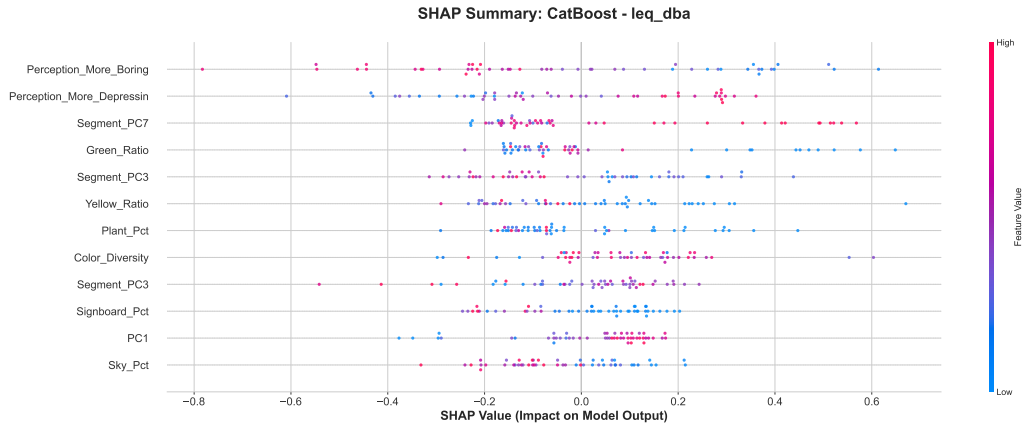


Figure A.17: SHAP Analysis CatBoost leq_dba Comprehensive 1

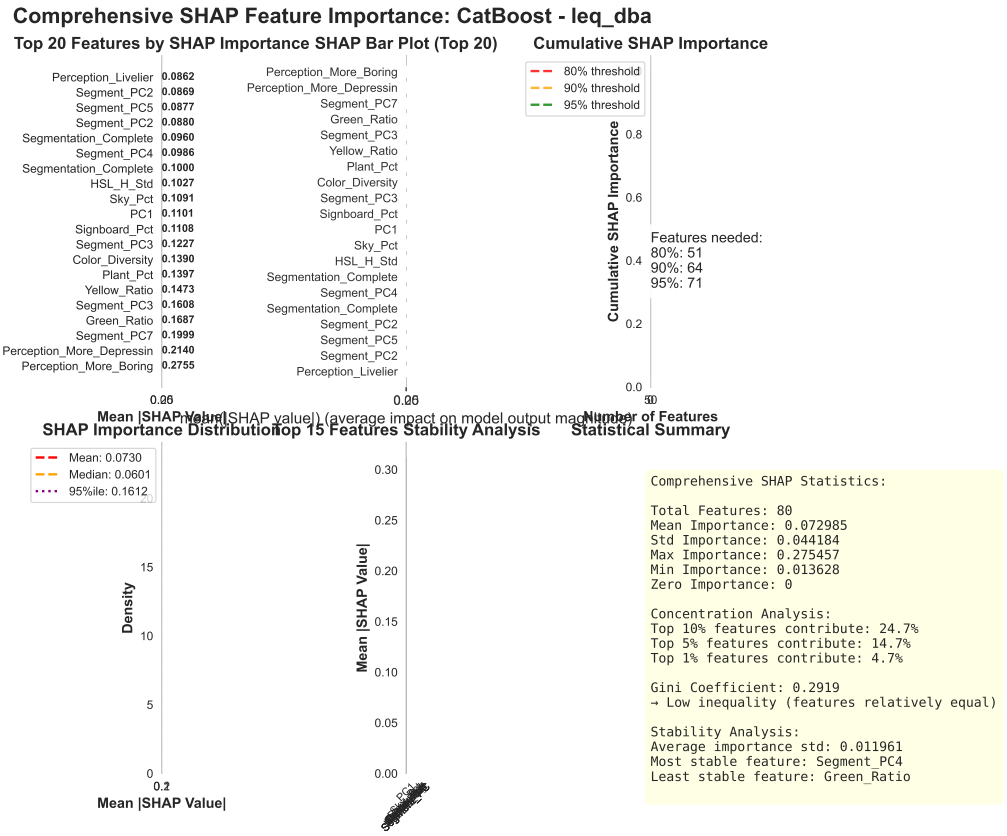


Figure A.18: SHAP Analysis CatBoost leq_dba Comprehensive 2

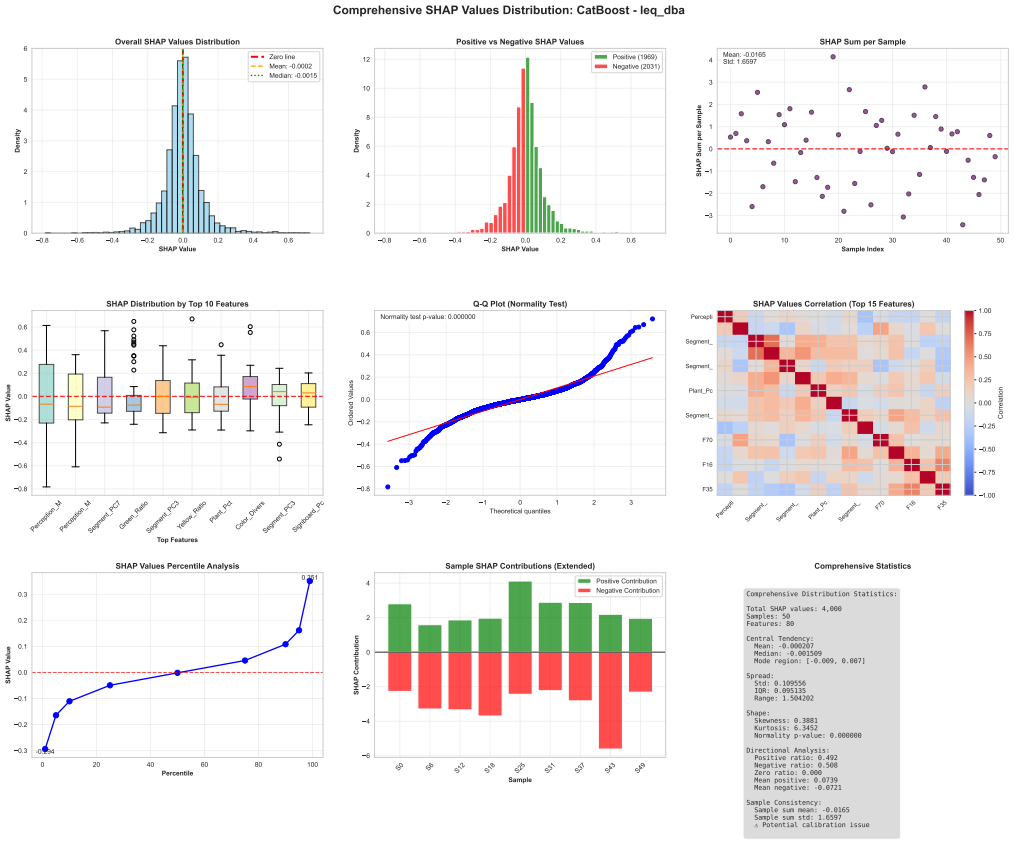


Figure A.19: SHAP Analysis CatBoost leq_dba Comprehensive 3

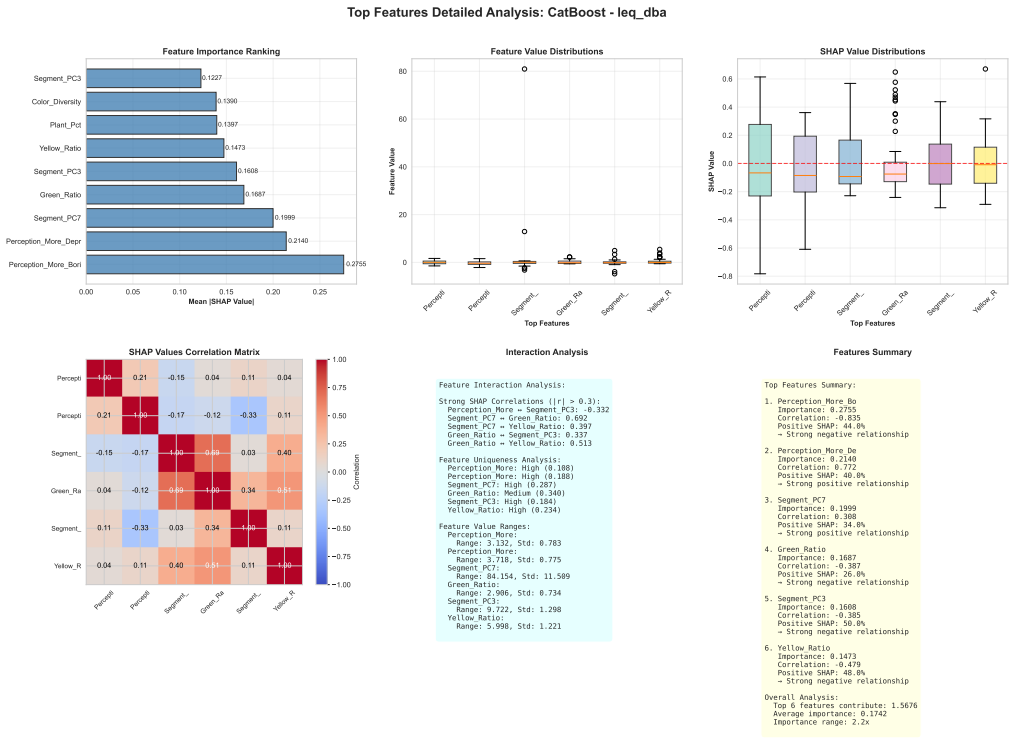


Figure A.20: SHAP Analysis CatBoost leq_dba Comprehensive 5

Figure A.21: SHAP Analysis CatBoost leq_dba Comprehensive 6

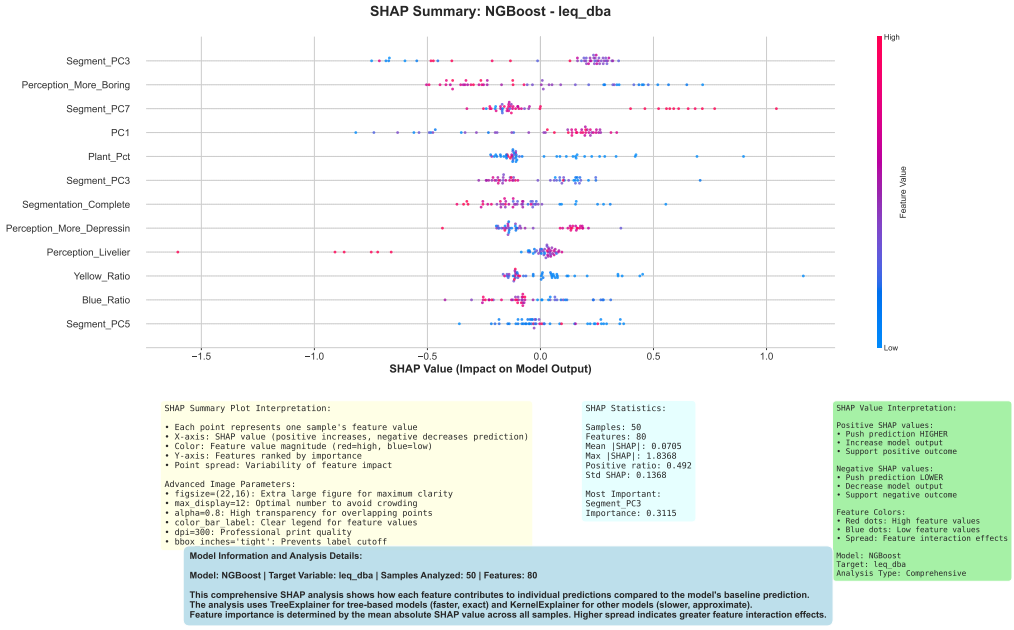


Figure A.22: SHAP Analysis NGBoost leq_dba Comprehensive 1

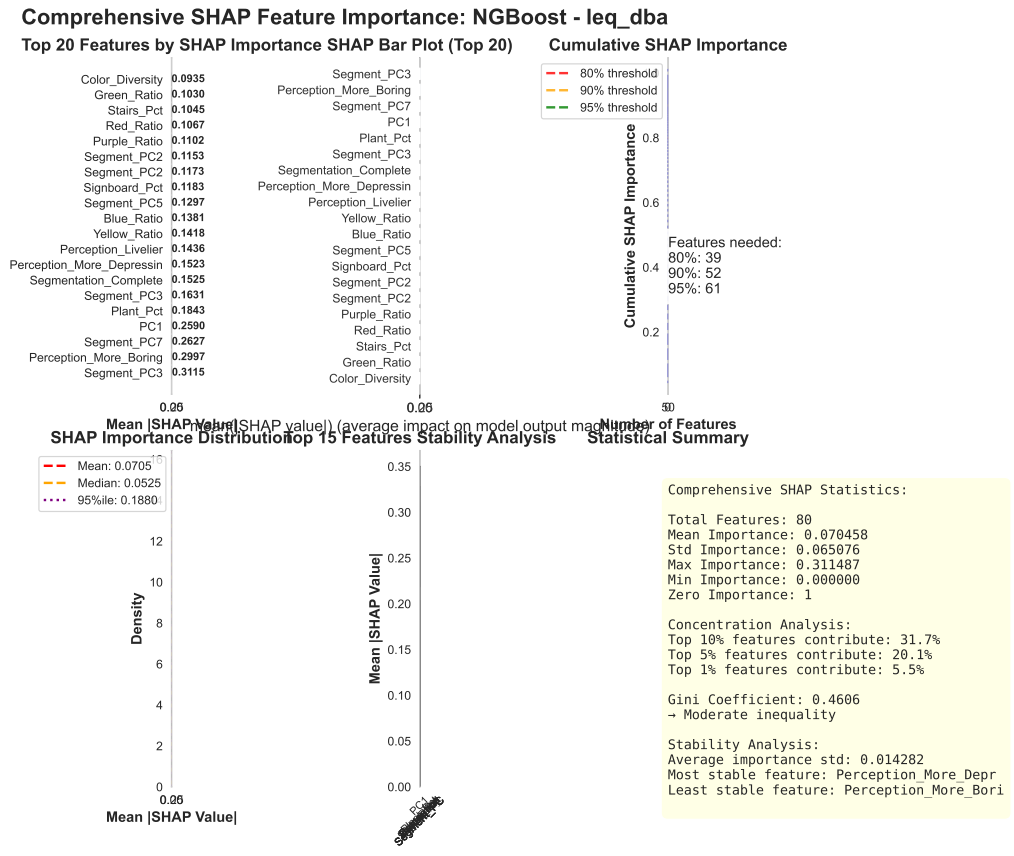


Figure A.23: SHAP Analysis NGBoost leq_dba Comprehensive 2

Figure A.24: SHAP Analysis NGBoost leq_dba Comprehensive 3

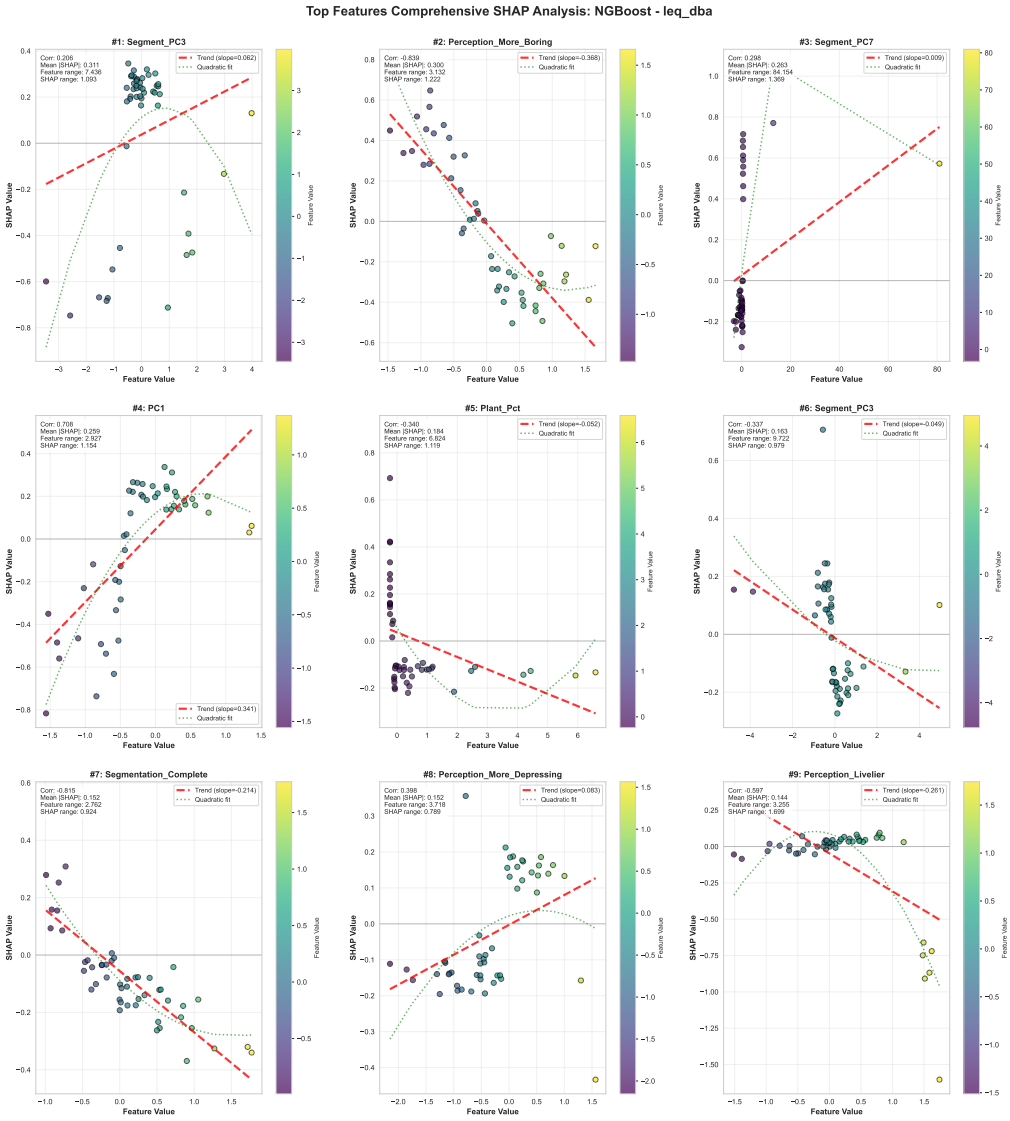


Figure A.25: SHAP Analysis NGBoost leq_dba Comprehensive 4

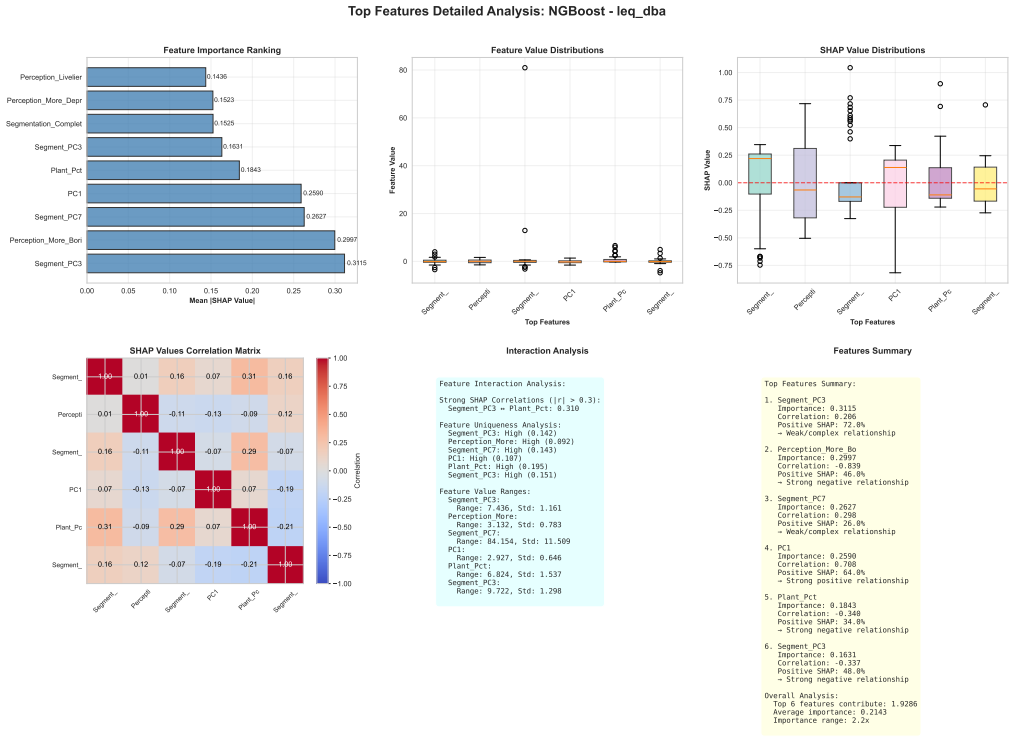


Figure A.26: SHAP Analysis NGBoost leq_dba Comprehensive 5

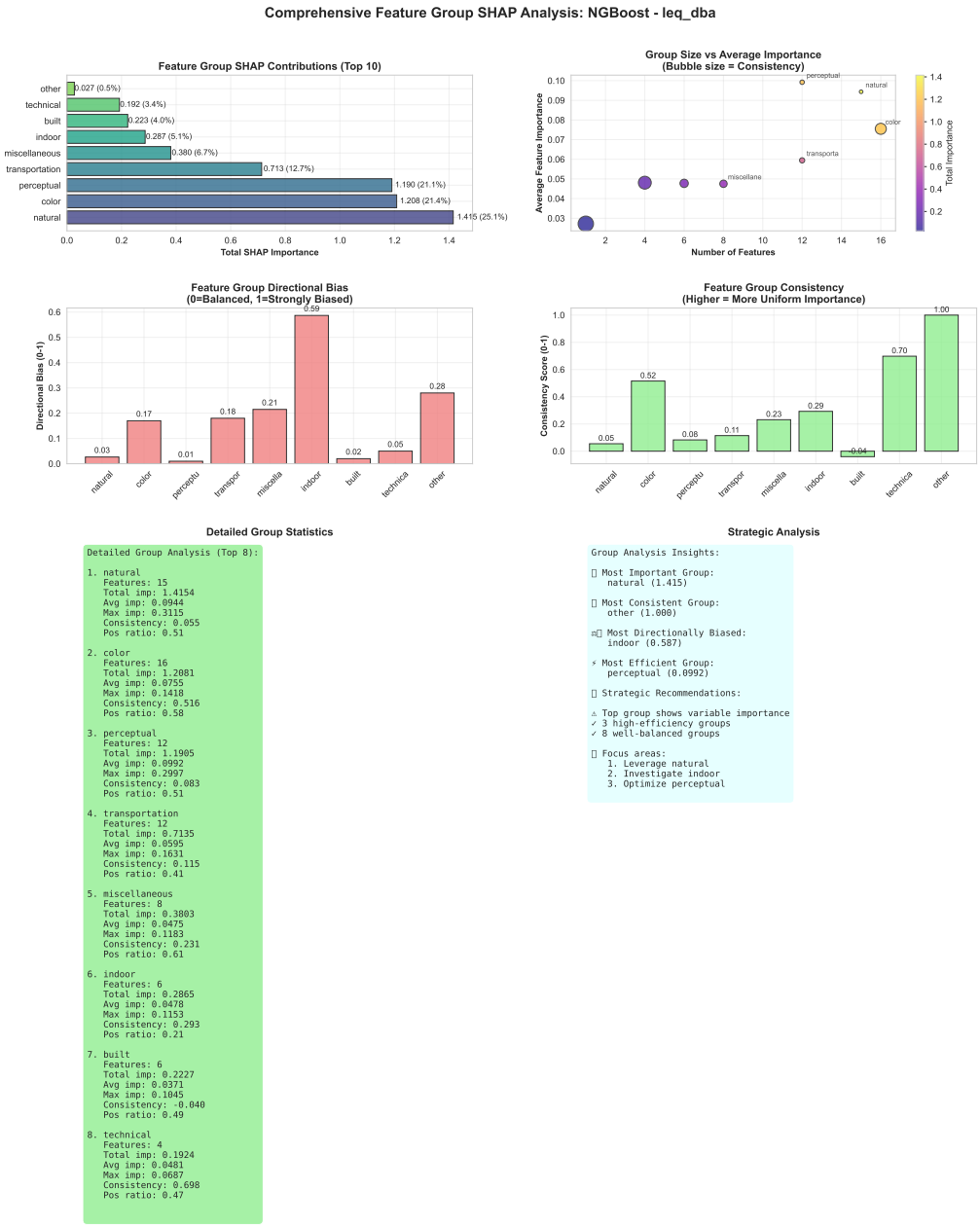


Figure A.27: SHAP Analysis NGBoost leq_dba Comprehensive 6

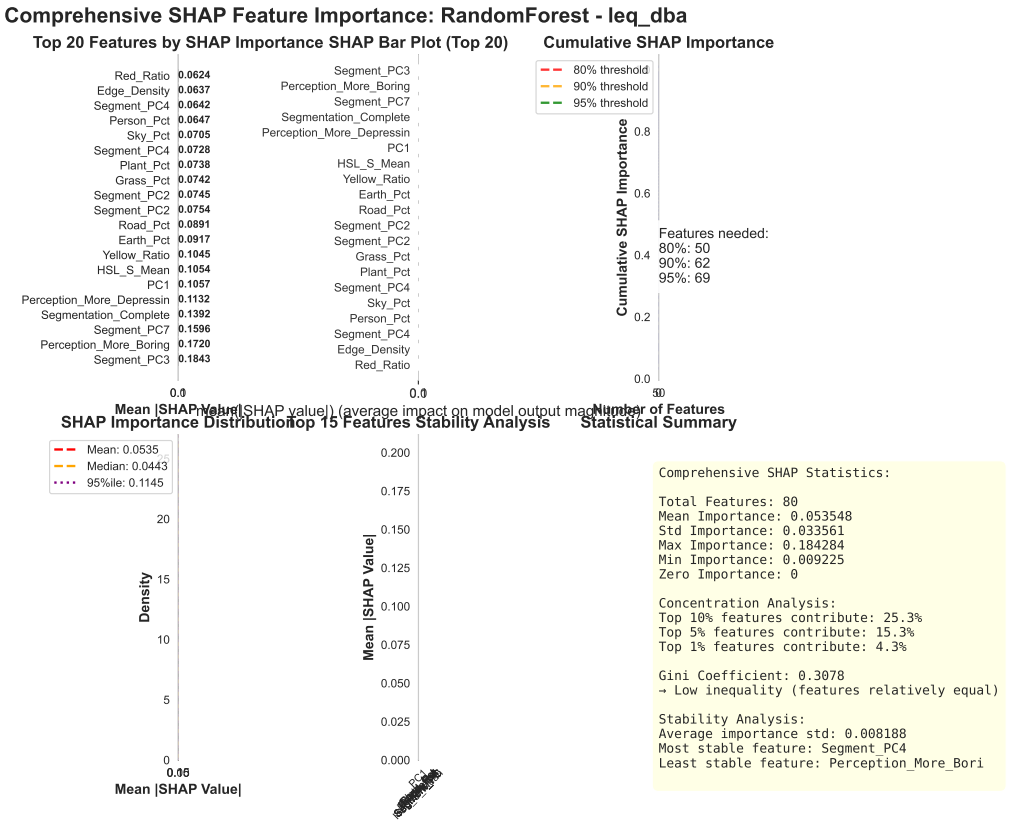


Figure A.28: SHAP Analysis RandomForest leq_dba Comprehensive 2

Figure A.29: SHAP Analysis RandomForest leq_dba Comprehensive 3

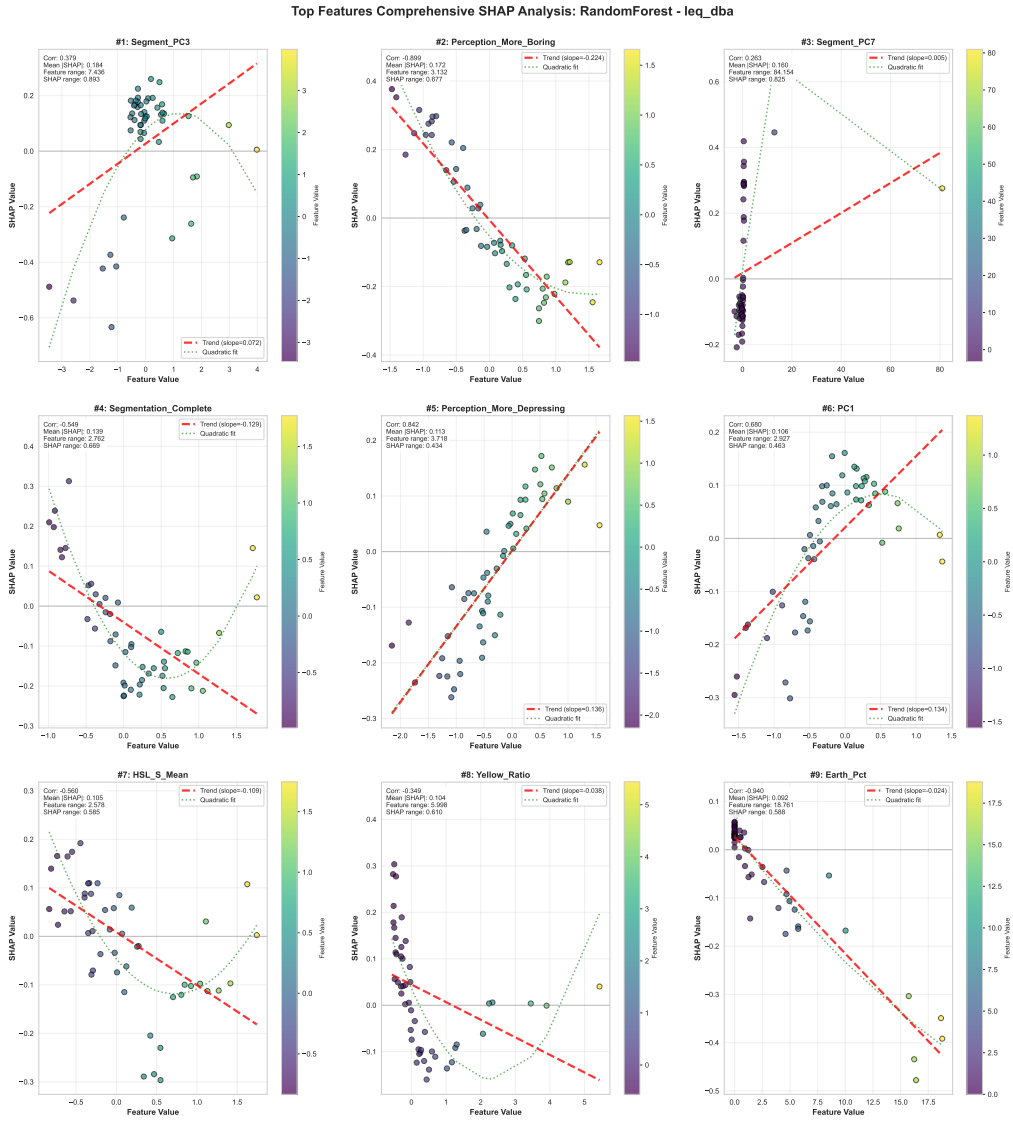


Figure A.30: SHAP Analysis RandomForest leq_dba Comprehensive 4

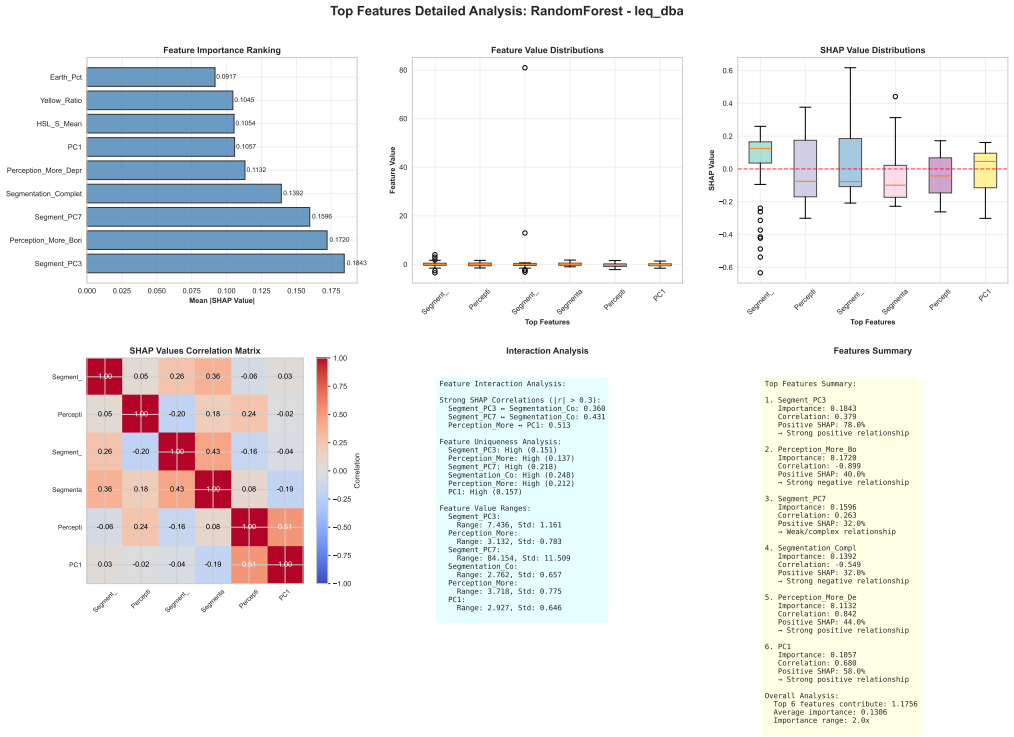


Figure A.31: SHAP Analysis RandomForest leq_dba Comprehensive 5

Figure A.32: SHAP Analysis RandomForest leq_dba Comprehensive 6

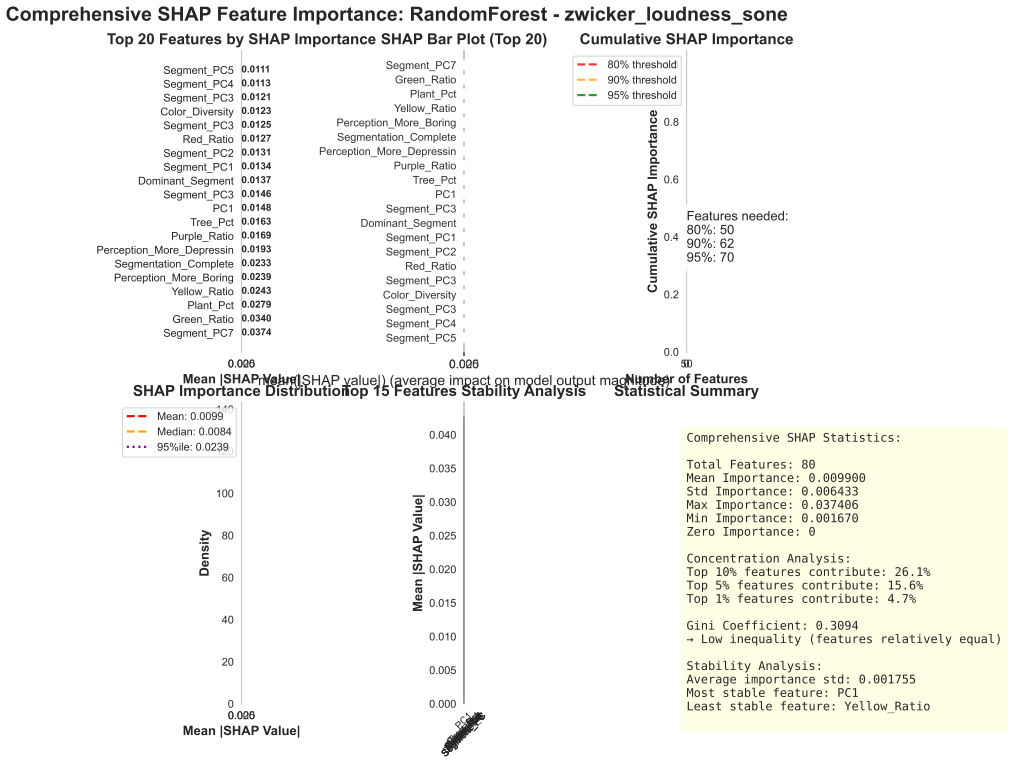


Figure A.33: SHAP Analysis RandomForest zwicker_loudness_sone Comprehensive 2

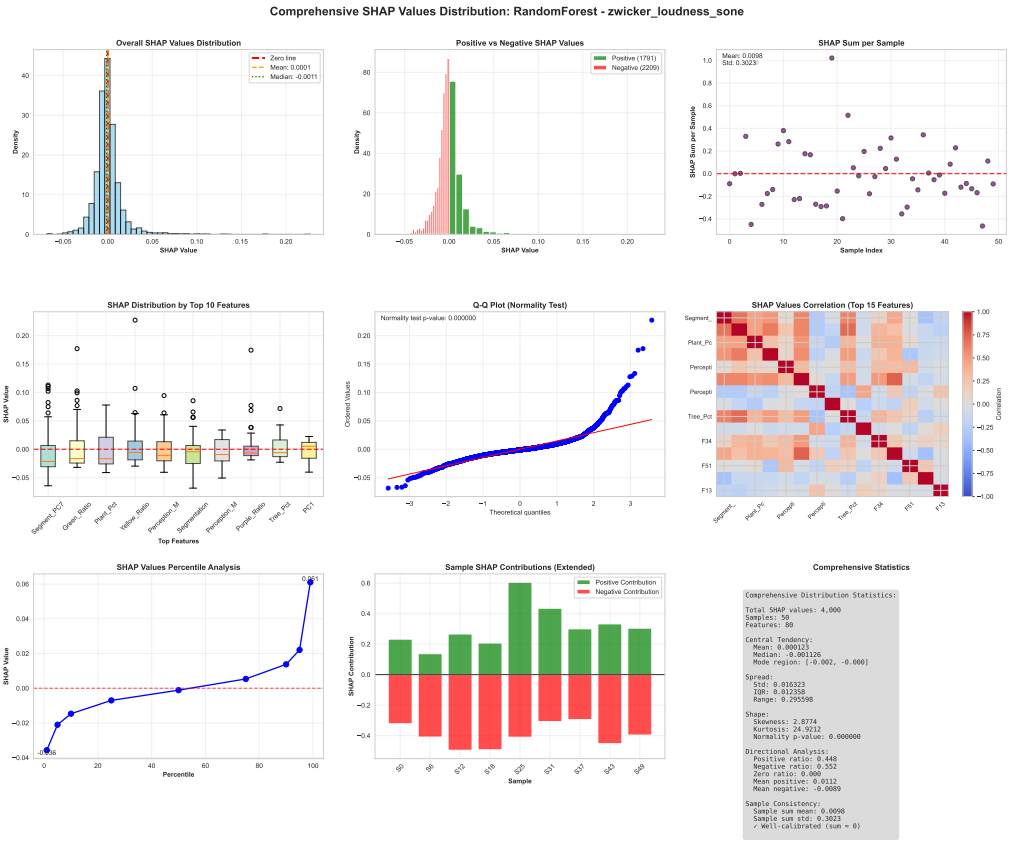


Figure A.34: SHAP Analysis RandomForest zwicker_loudness_sone Comprehensive 3

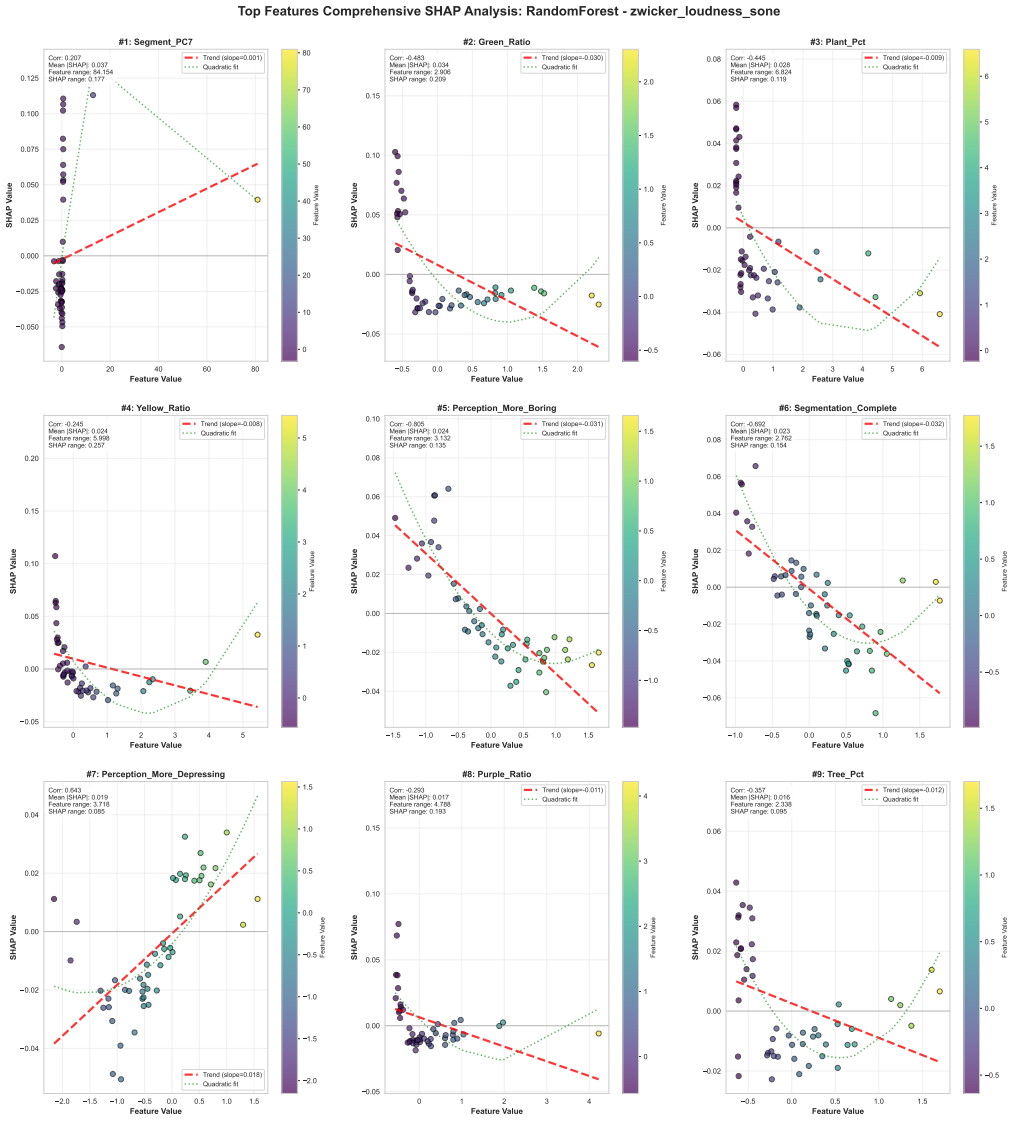


Figure A.35: SHAP Analysis RandomForest zwicker_loudness_sone Comprehensive 4

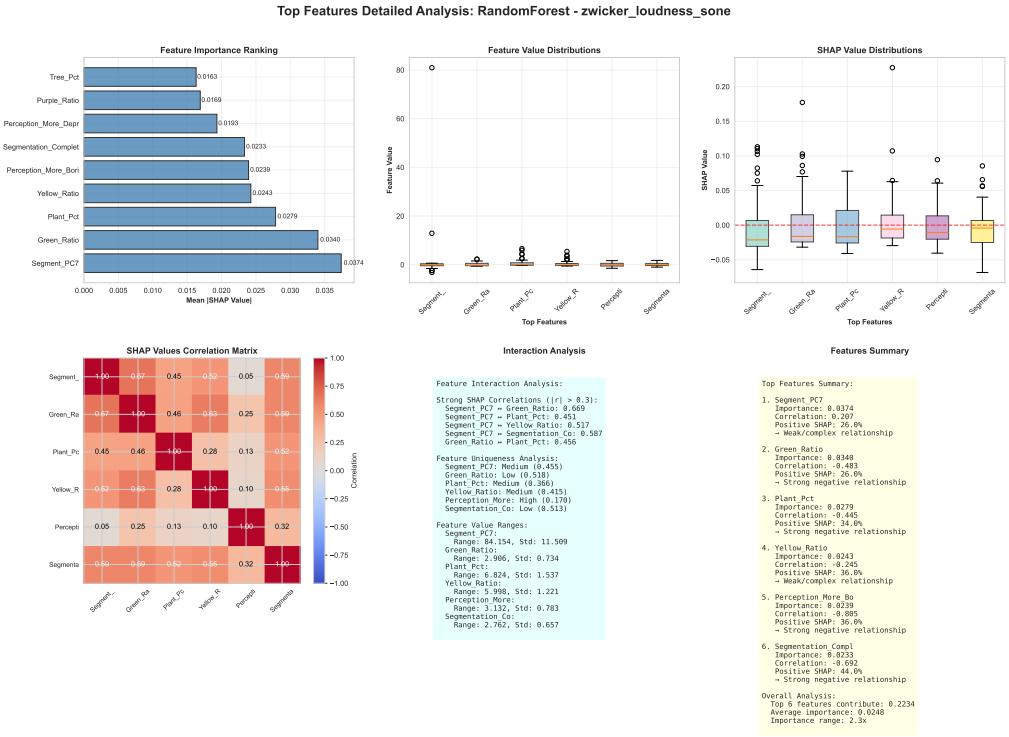


Figure A.36: SHAP Analysis RandomForest zwicker_loudness_sone Comprehensive 5

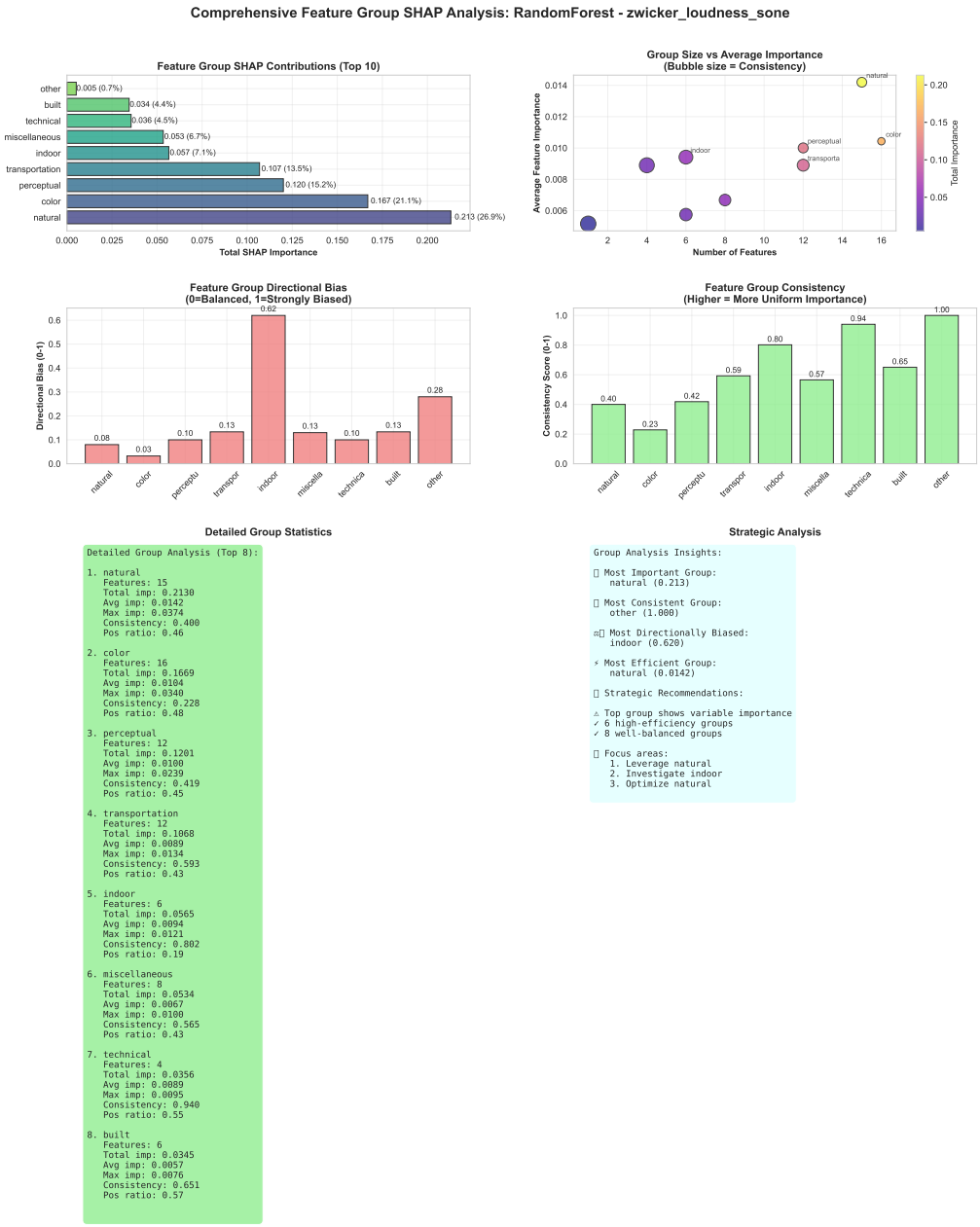


Figure A.37: SHAP Analysis RandomForest zwicker_loudness_sone Comprehensive 6

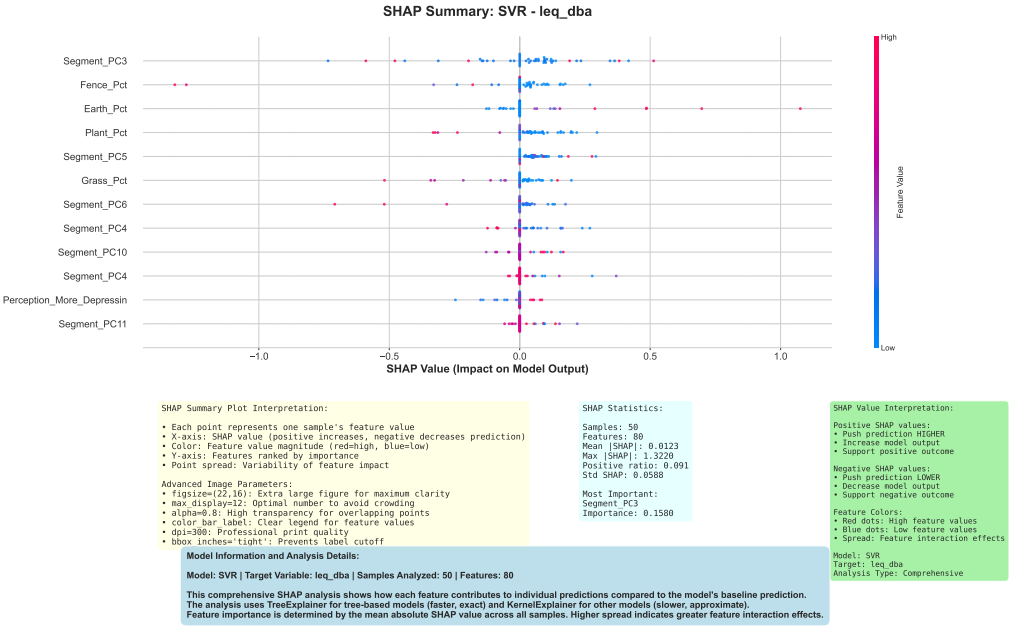


Figure A.38: SHAP Analysis SVR leq_dba Comprehensive 1

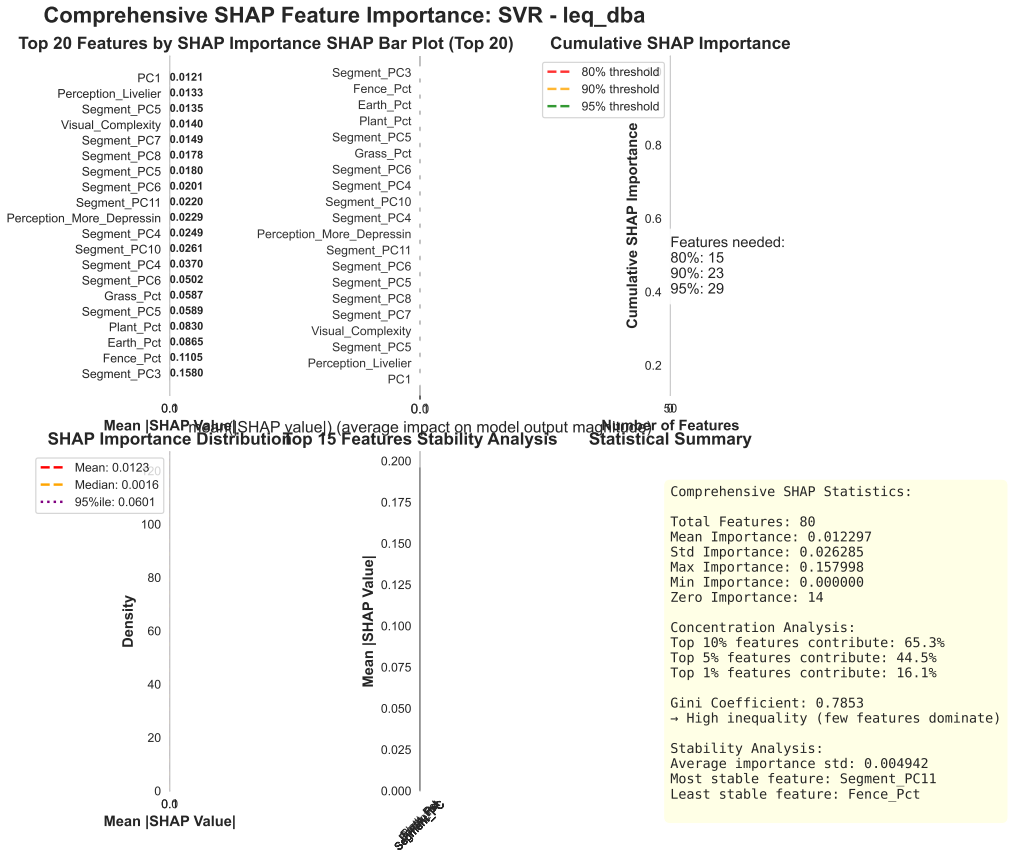


Figure A.39: SHAP Analysis SVR leq_dba Comprehensive 2

Figure A.40: SHAP Analysis SVR leq_dba Comprehensive 3

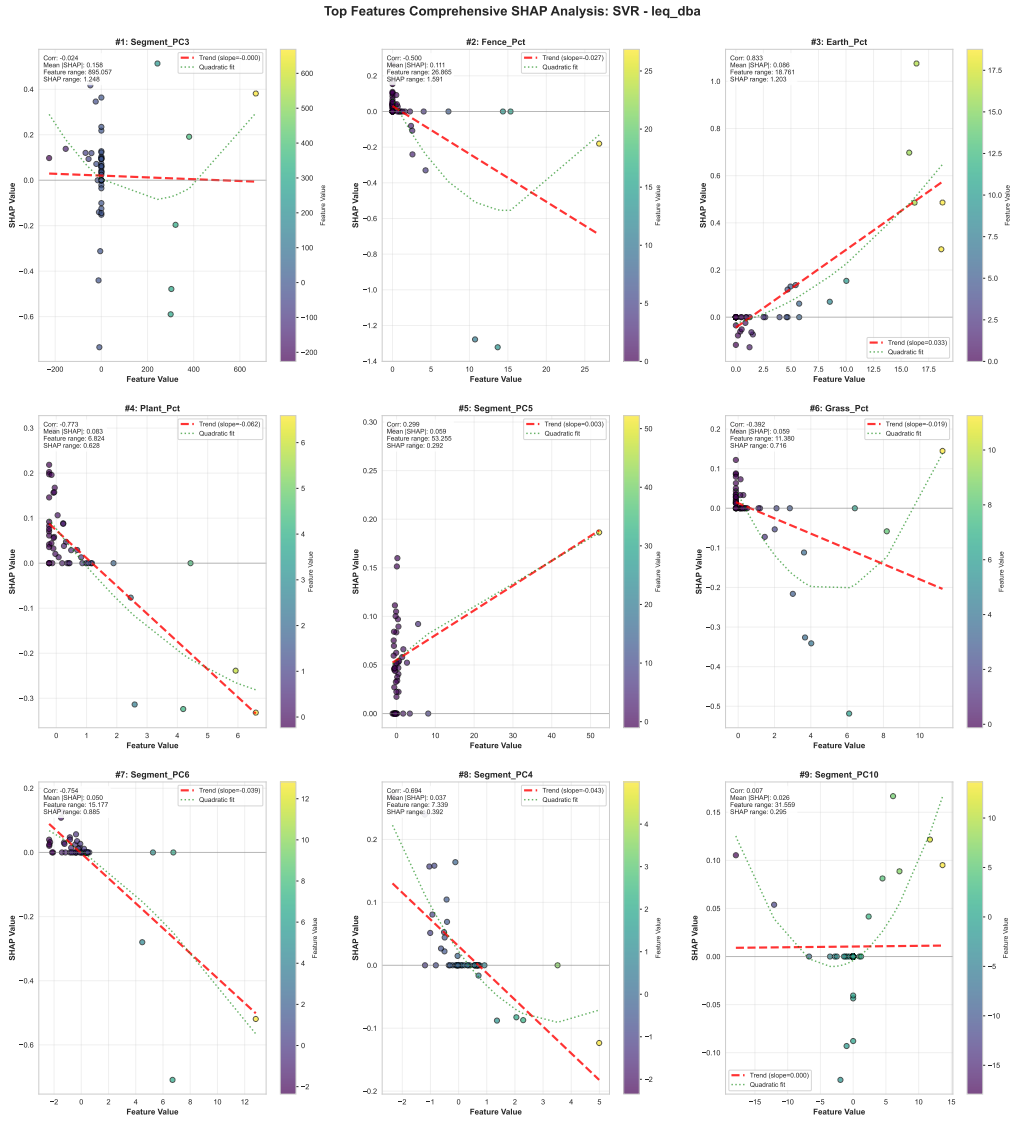


Figure A.41: SHAP Analysis SVR leq_dba Comprehensive 4

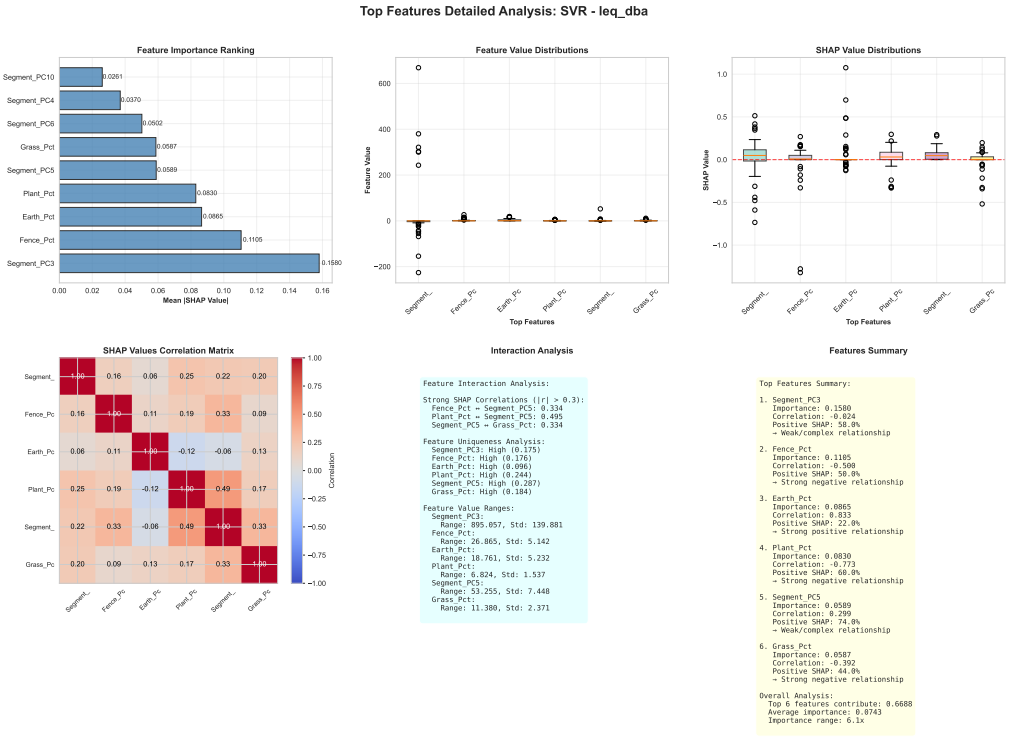


Figure A.42: SHAP Analysis SVR leq_dba Comprehensive 5

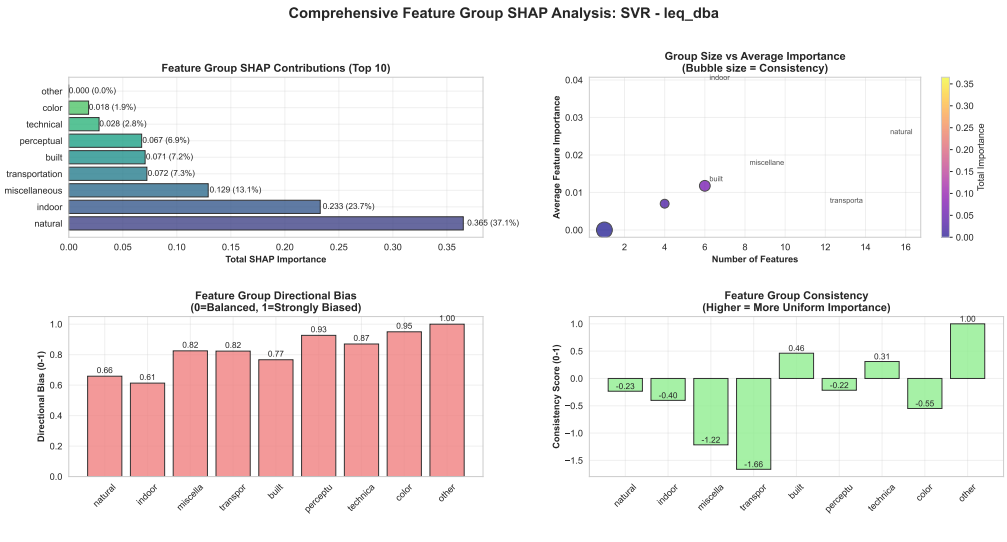


Figure A.43: SHAP Analysis SVR leq_dba Comprehensive 6

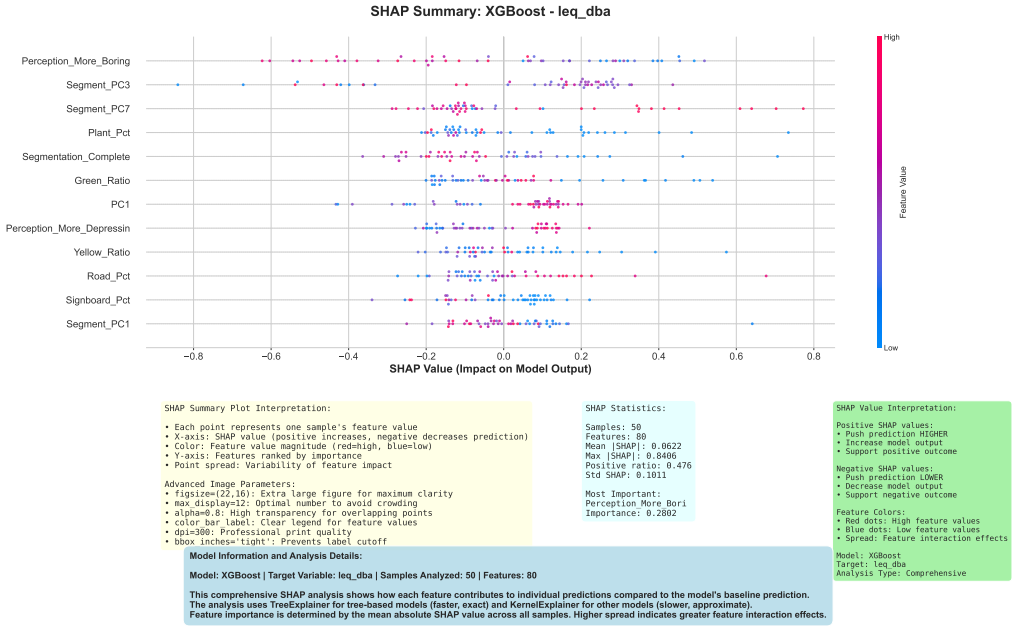


Figure A.44: SHAP Analysis XGBoost leq_dba Comprehensive 1

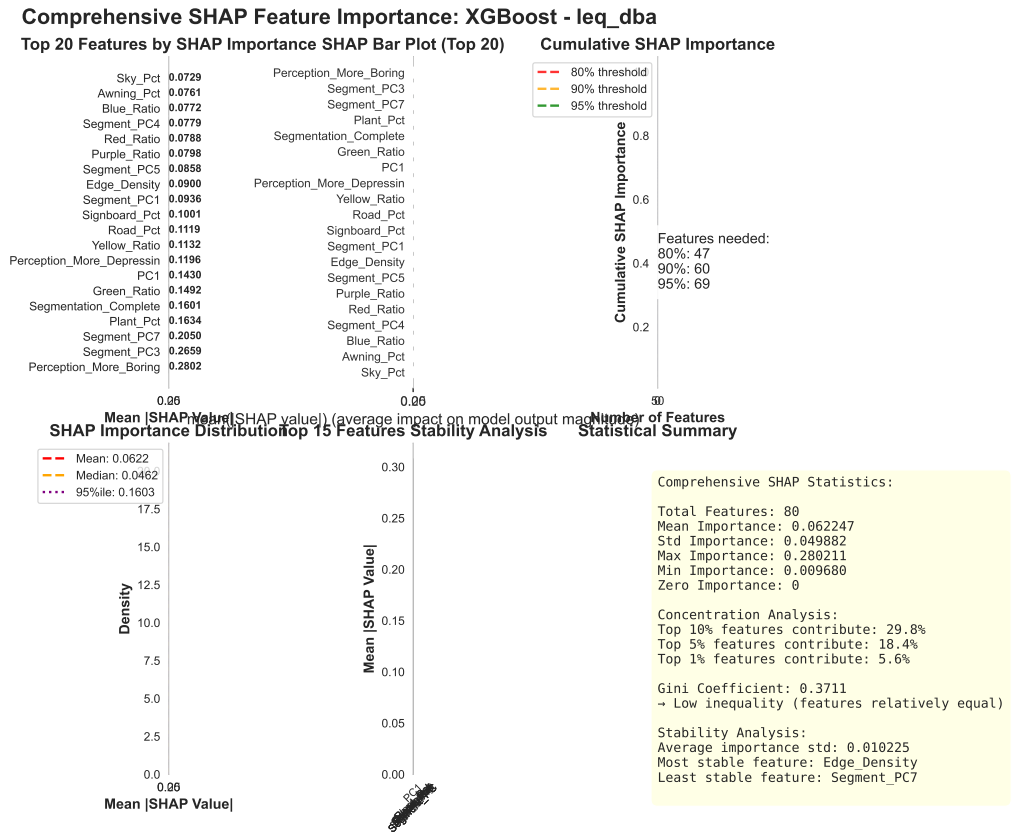


Figure A.45: SHAP Analysis XGBoost leq_dba Comprehensive 2

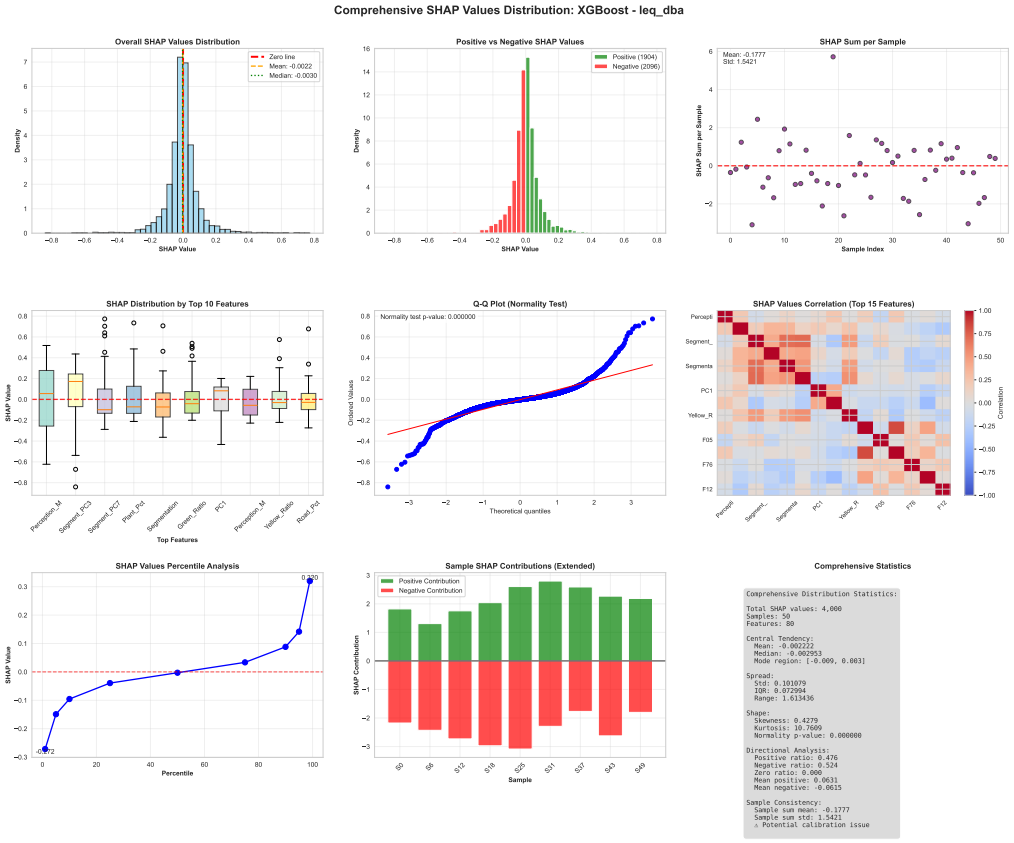


Figure A.46: SHAP Analysis XGBoost leq_dba Comprehensive 3

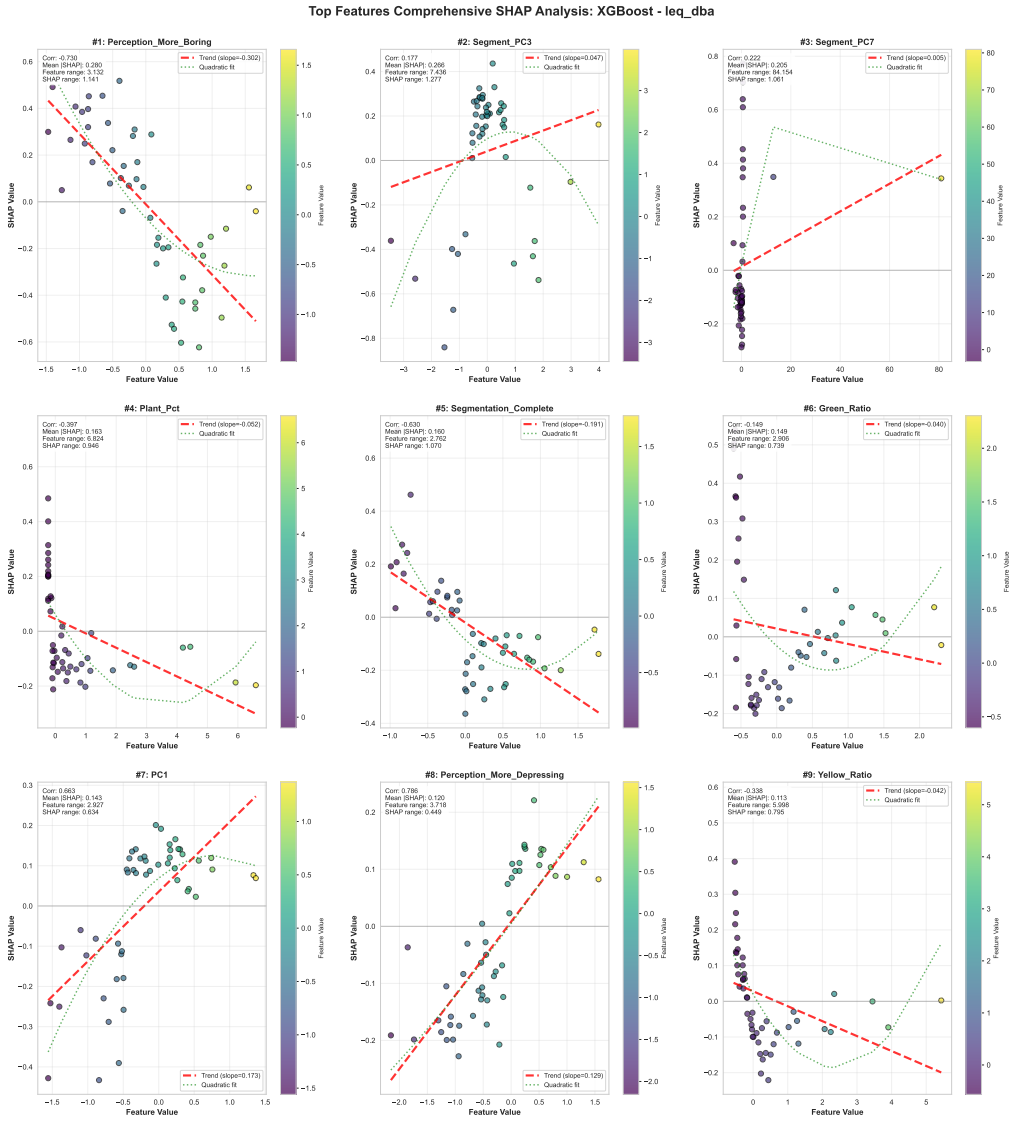


Figure A.47: SHAP Analysis XGBoost leq_dba Comprehensive 4

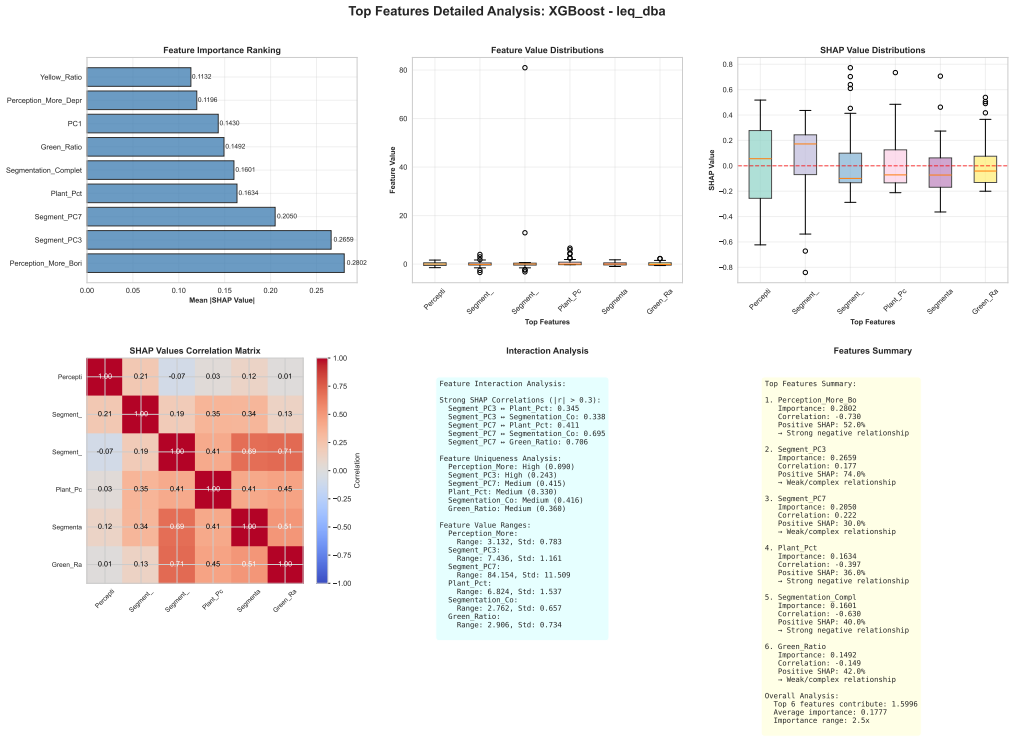


Figure A.48: SHAP Analysis XGBoost leq_dba Comprehensive 5

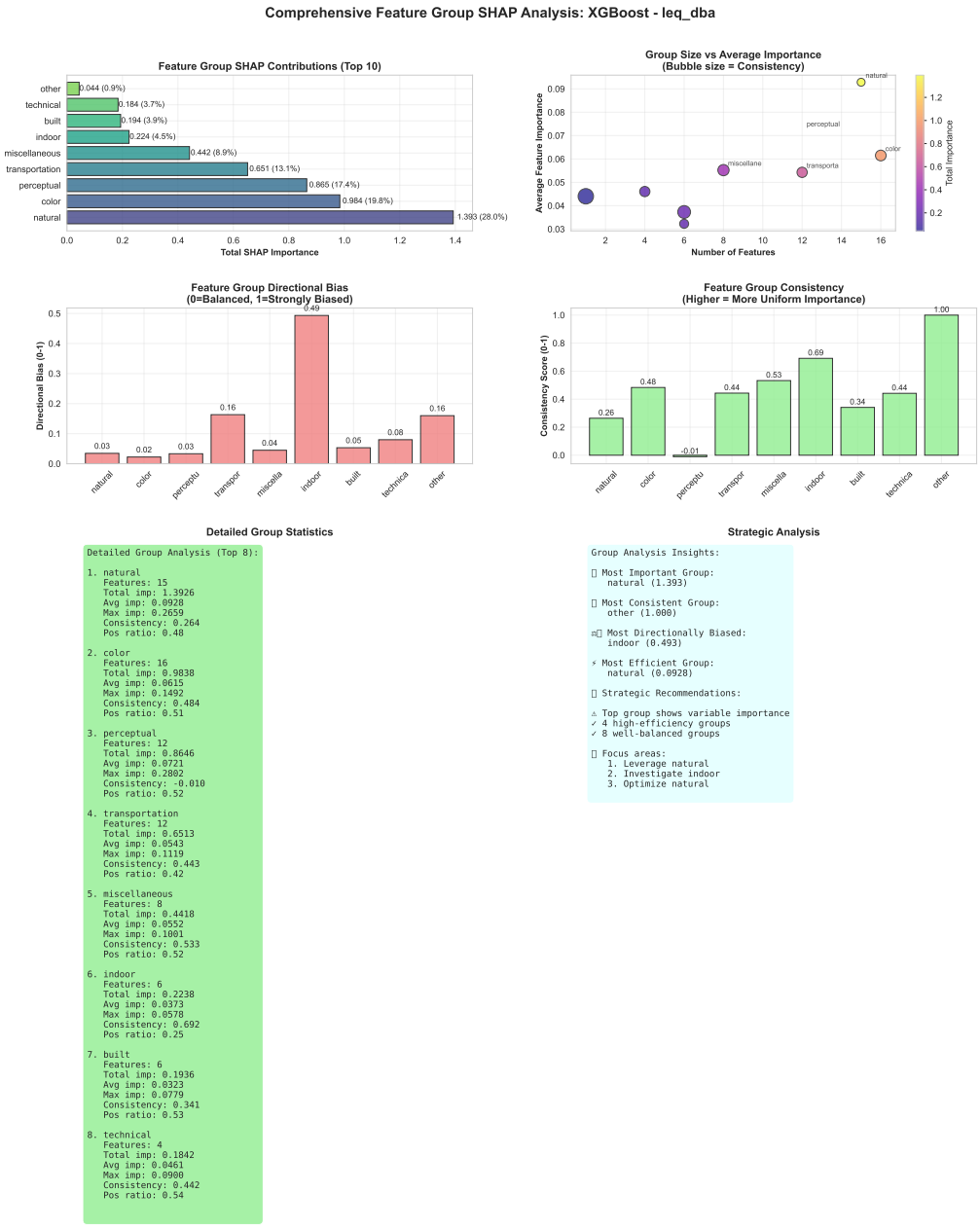


Figure A.49: SHAP Analysis XGBoost leq_dba Comprehensive 6

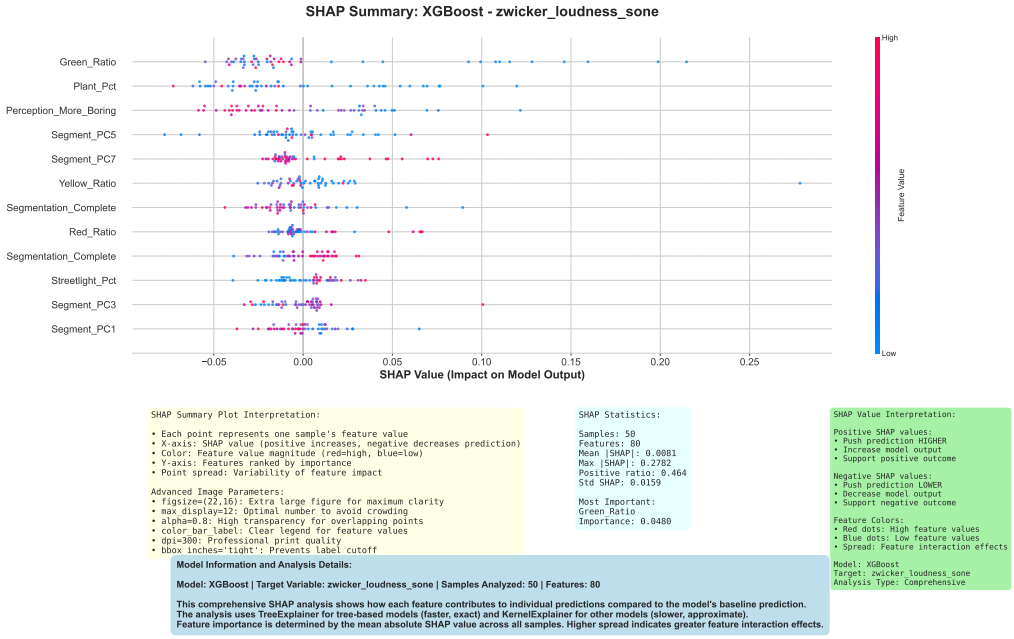


Figure A.50: SHAP Analysis XGBoost zwicker_loudness_sone Comprehensive 1

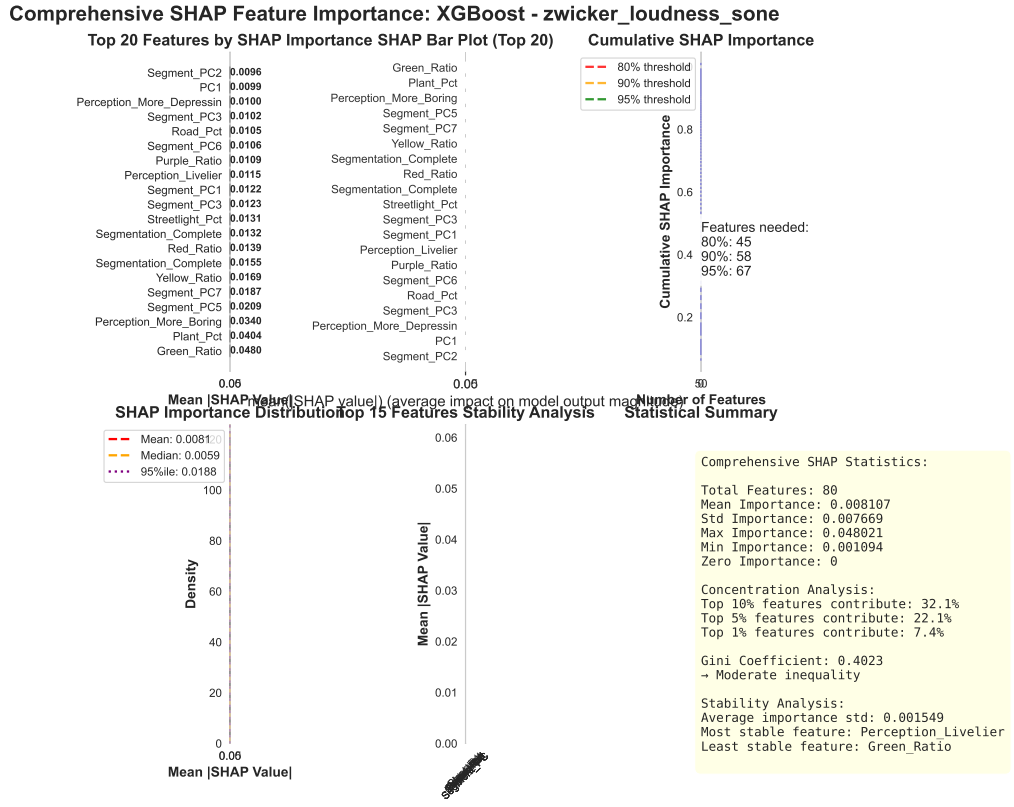


Figure A.51: SHAP Analysis XGBoost zwicker_loudness_sone Comprehensive 2

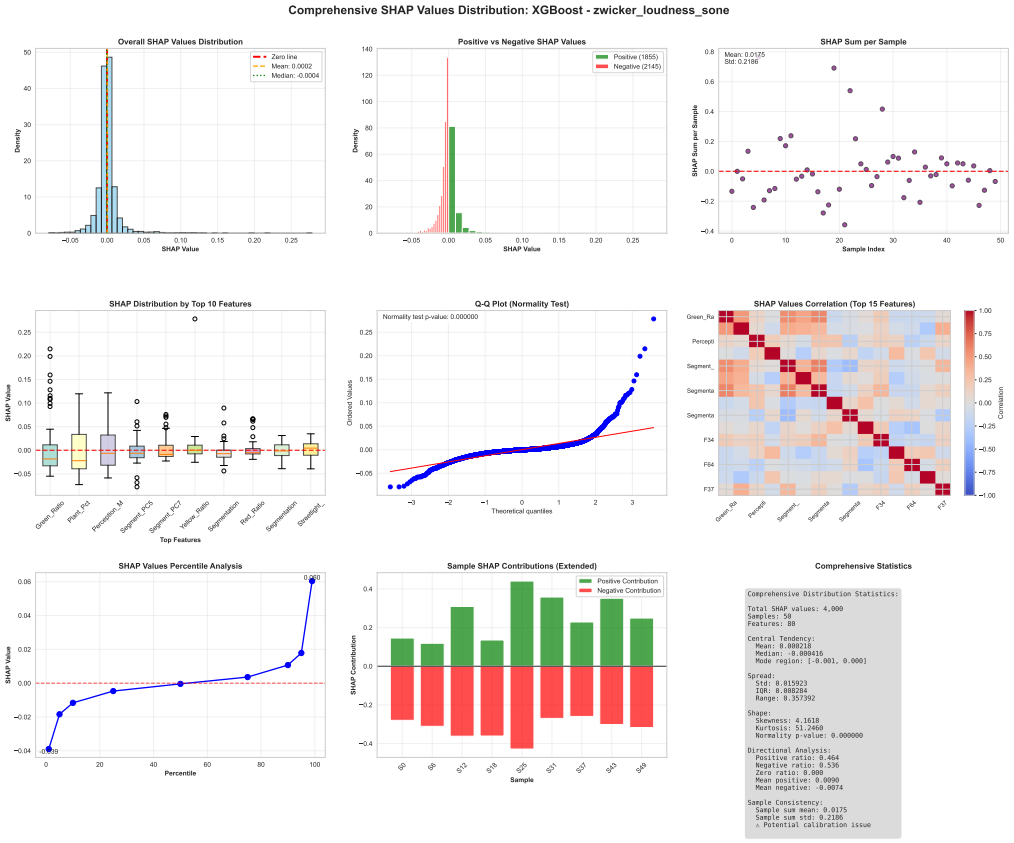


Figure A.52: SHAP Analysis XGBoost zwicker_loudness_sone Comprehensive 3

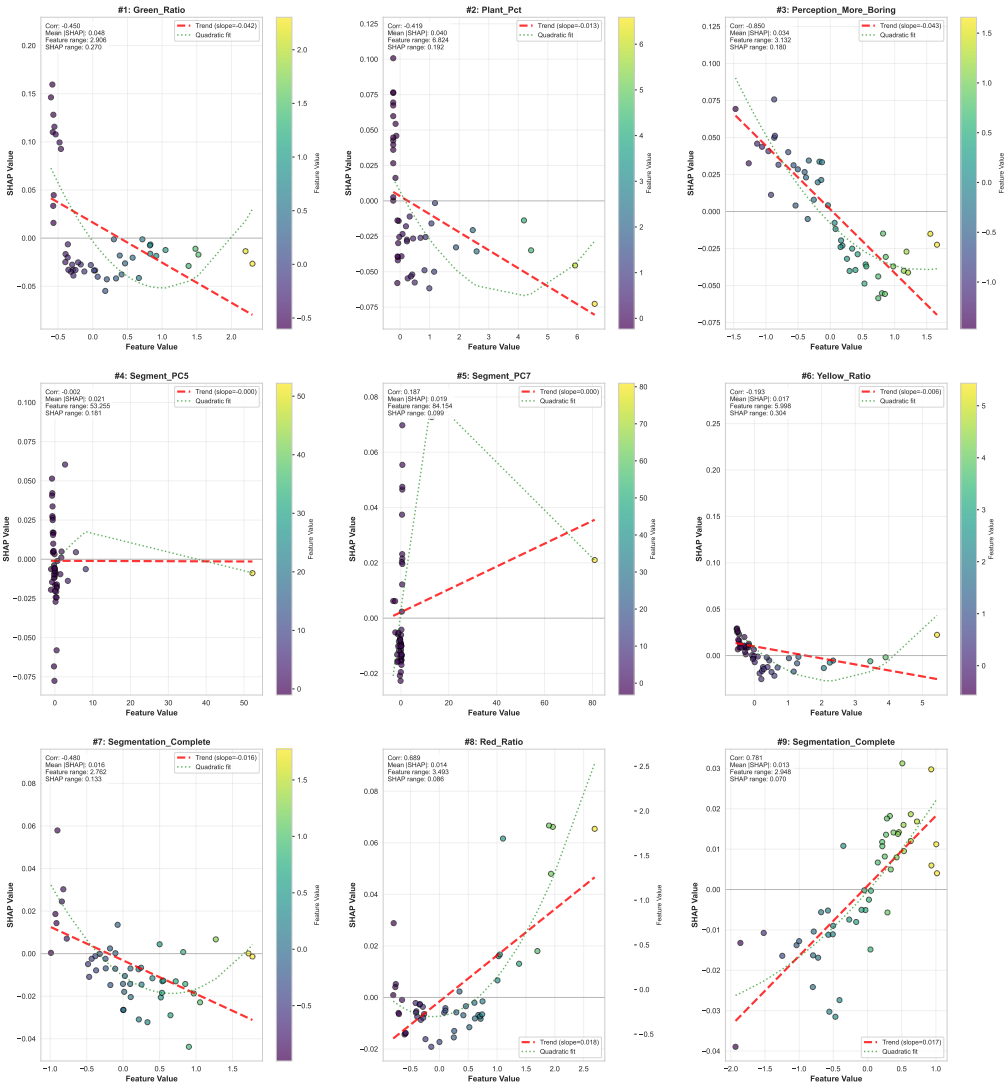


Figure A.53: SHAP Analysis XGBoost zwicker_loudness_sone Comprehensive 4

A.5. Residual Analysis

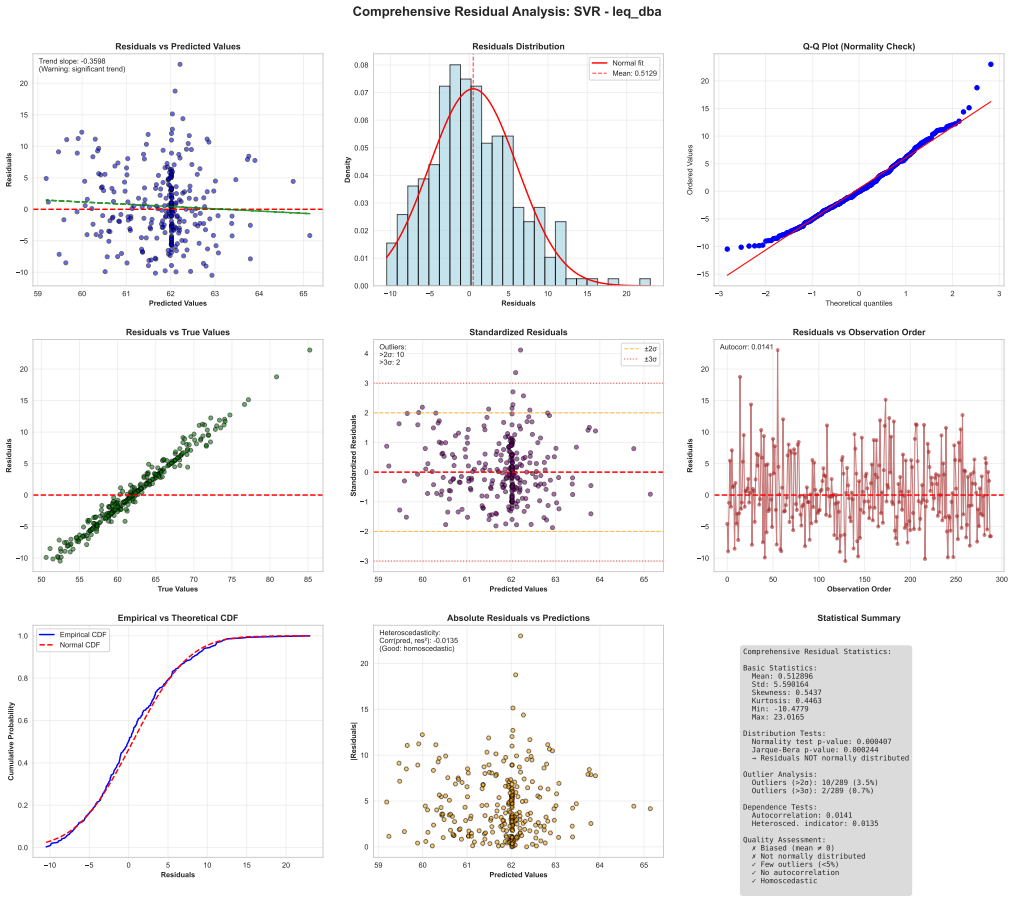


Figure A.54: Comprehensive Residual Analysis 1

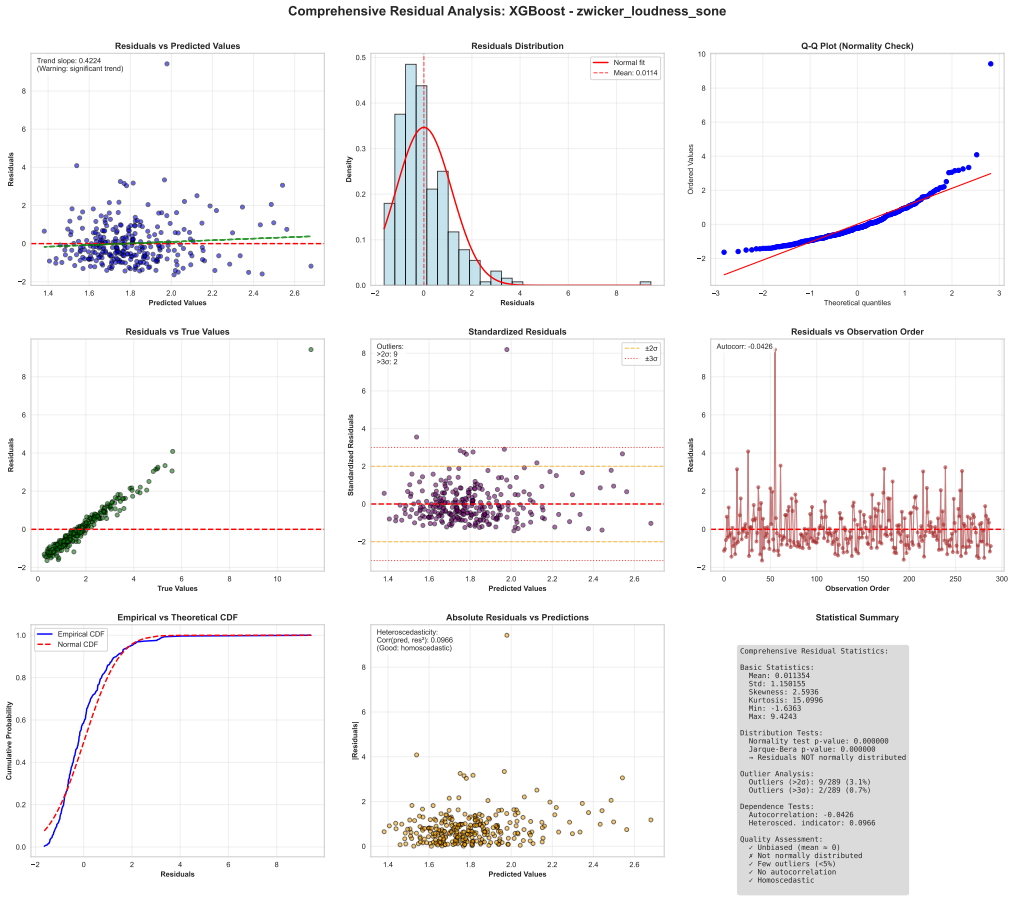


Figure A.55: Comprehensive Residual Analysis 10

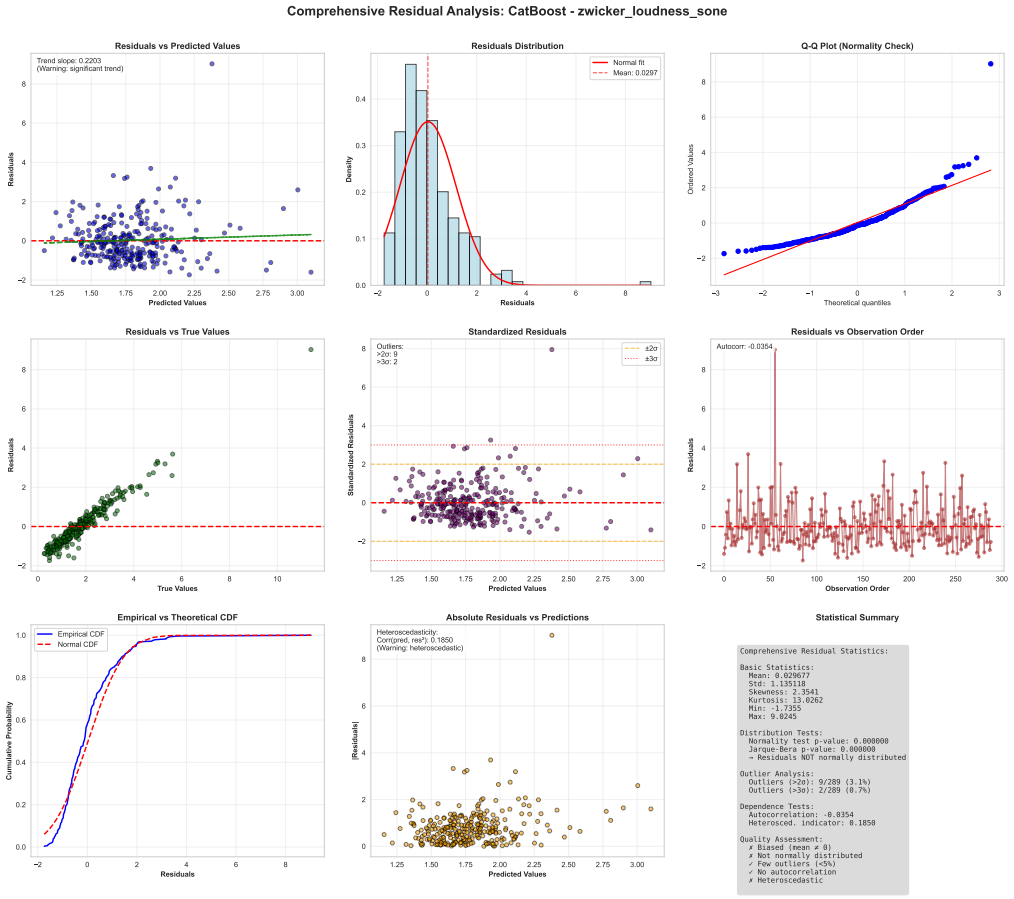


Figure A.56: Comprehensive Residual Analysis 11

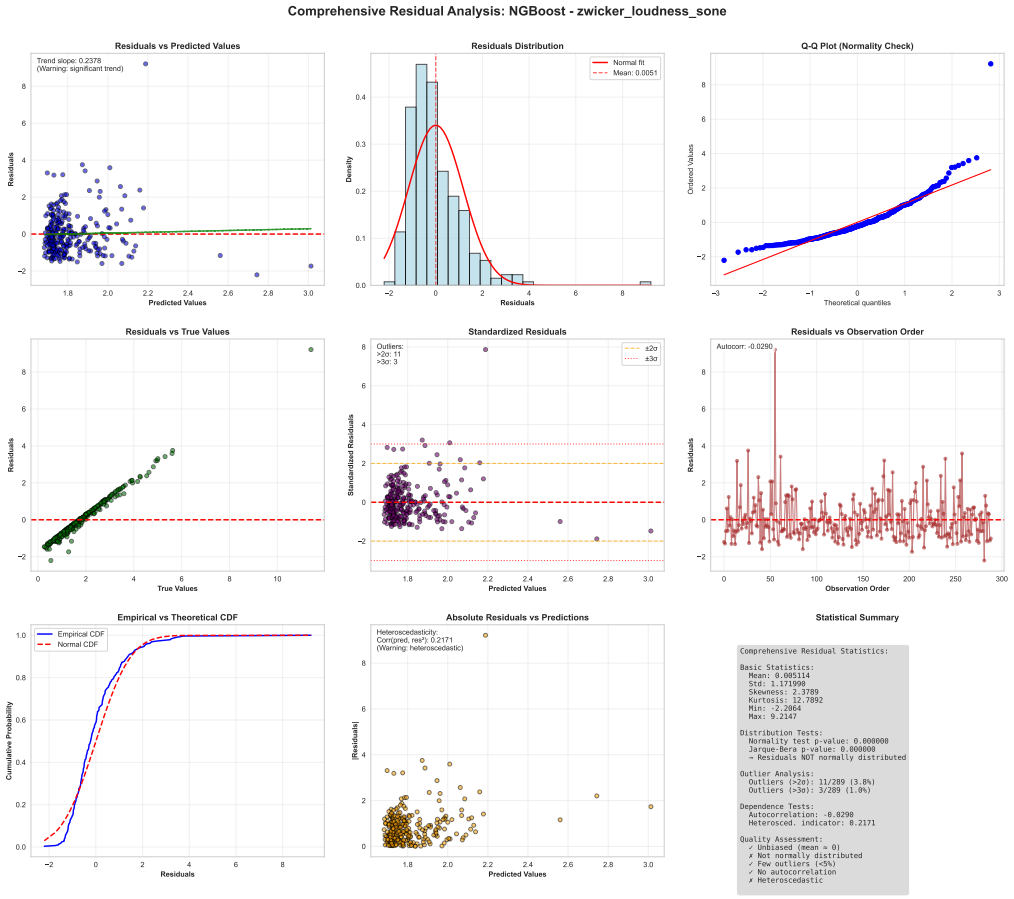


Figure A.57: Comprehensive Residual Analysis 12

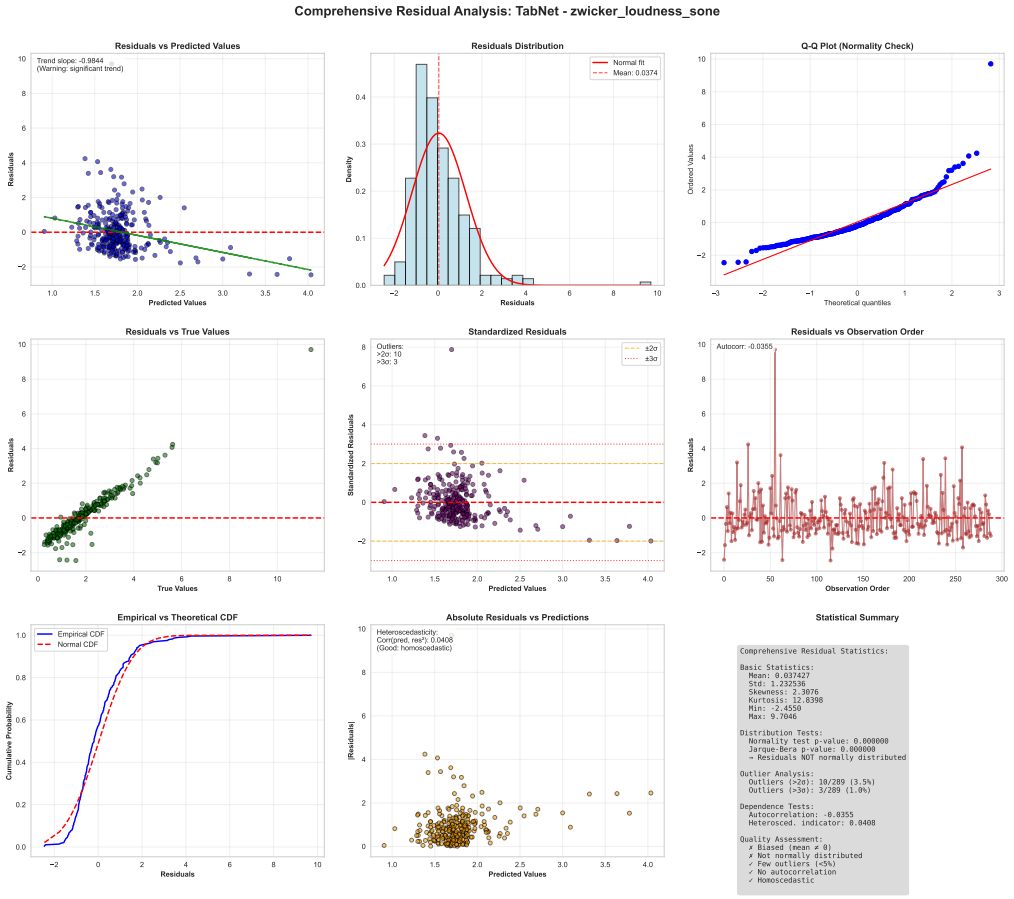


Figure A.58: Comprehensive Residual Analysis 13

Figure A.59: Comprehensive Residual Analysis 14

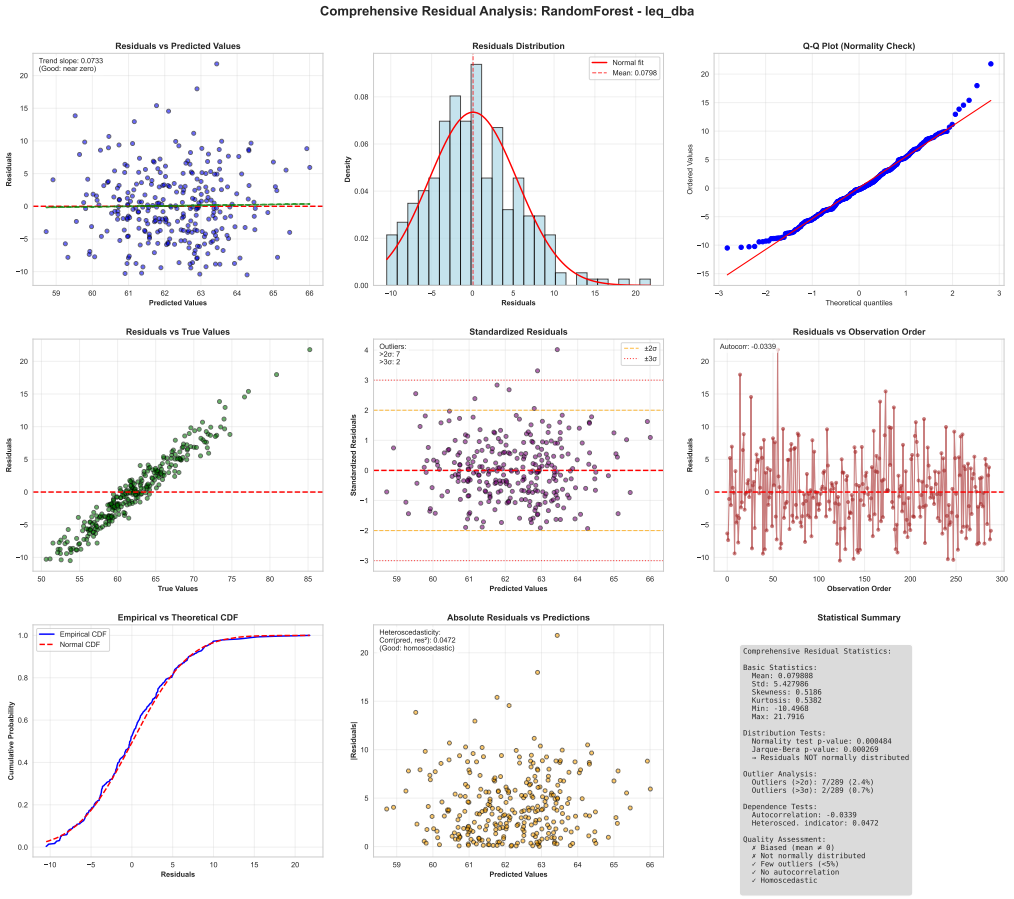


Figure A.60: Comprehensive Residual Analysis 2

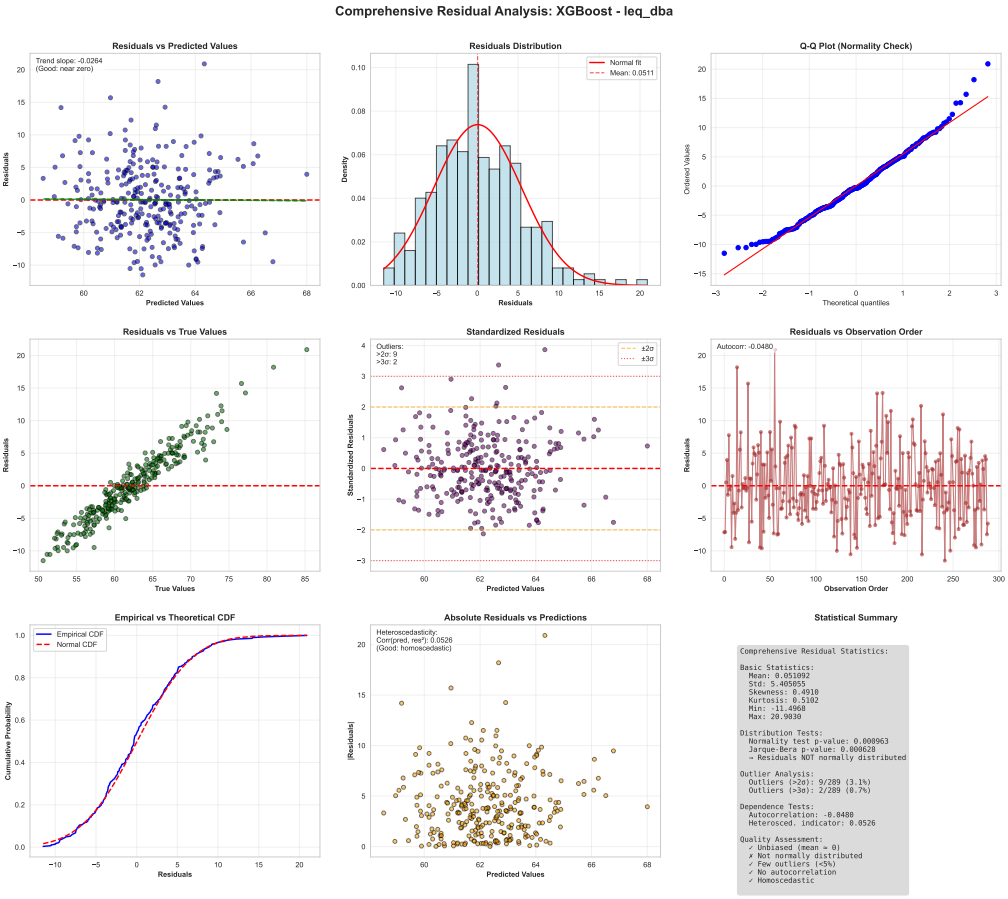


Figure A.61: Comprehensive Residual Analysis 3

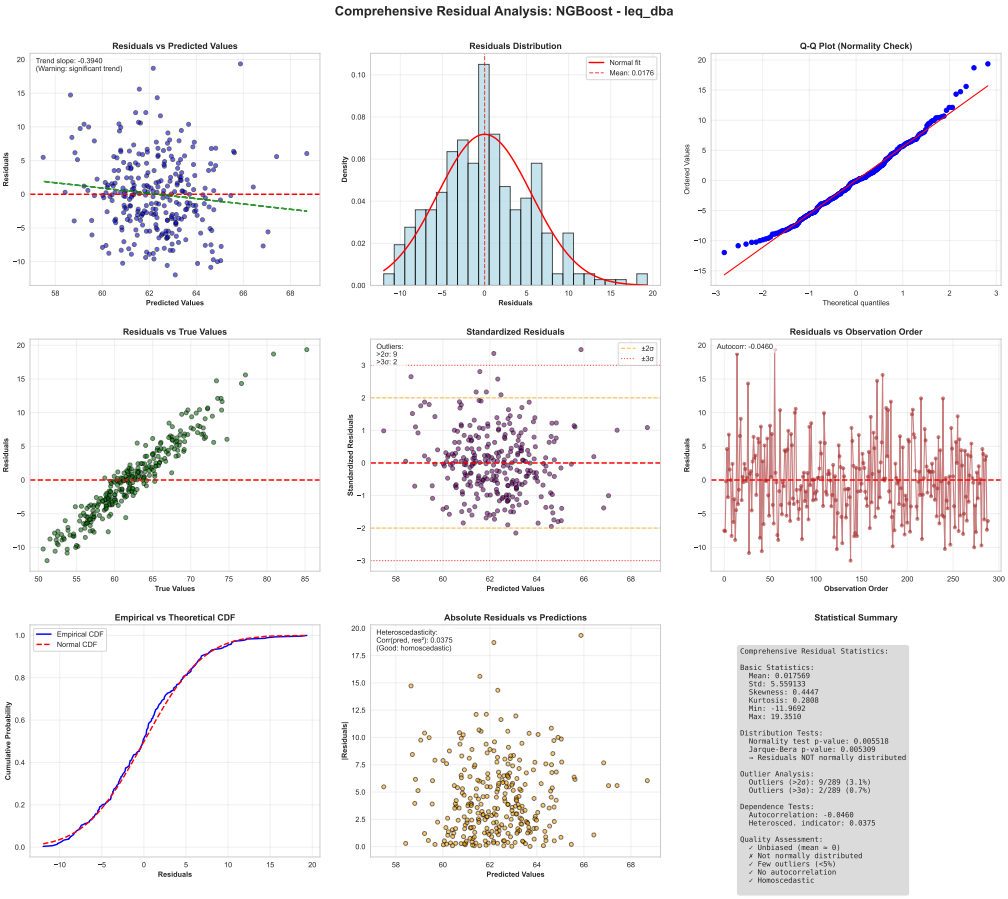


Figure A.62: Comprehensive Residual Analysis 5

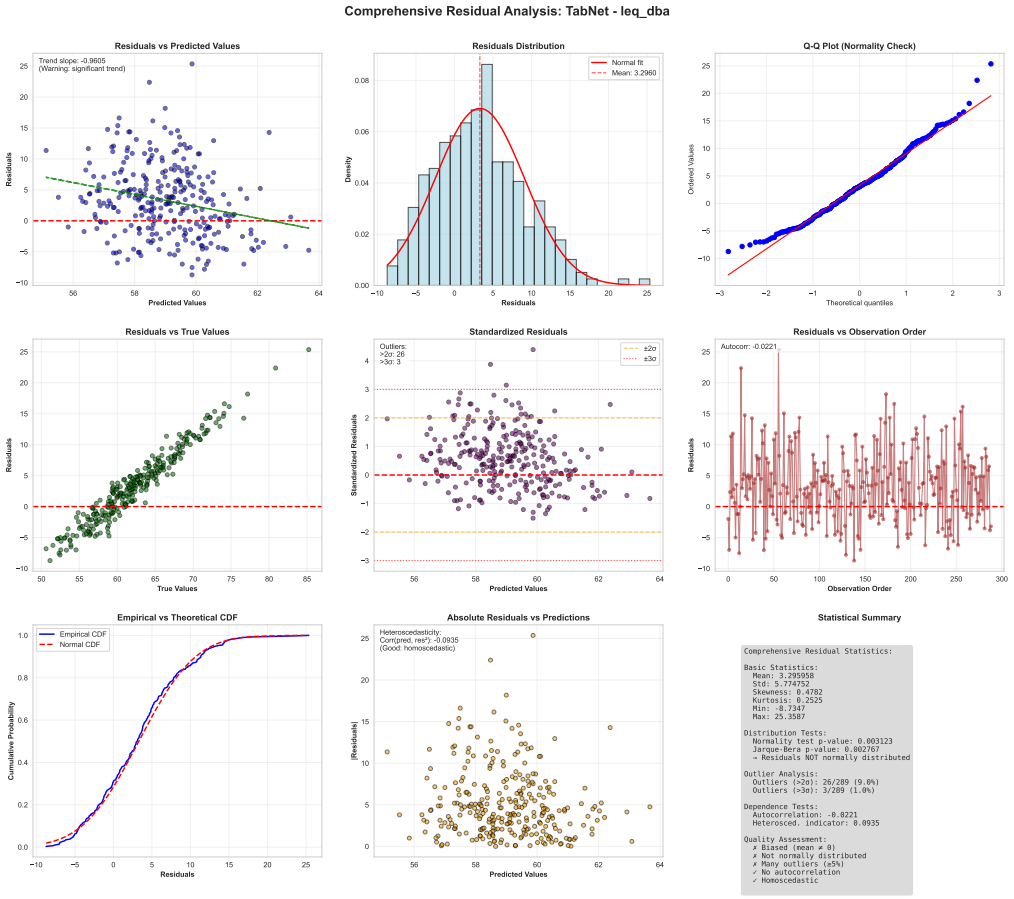


Figure A.63: Comprehensive Residual Analysis 6

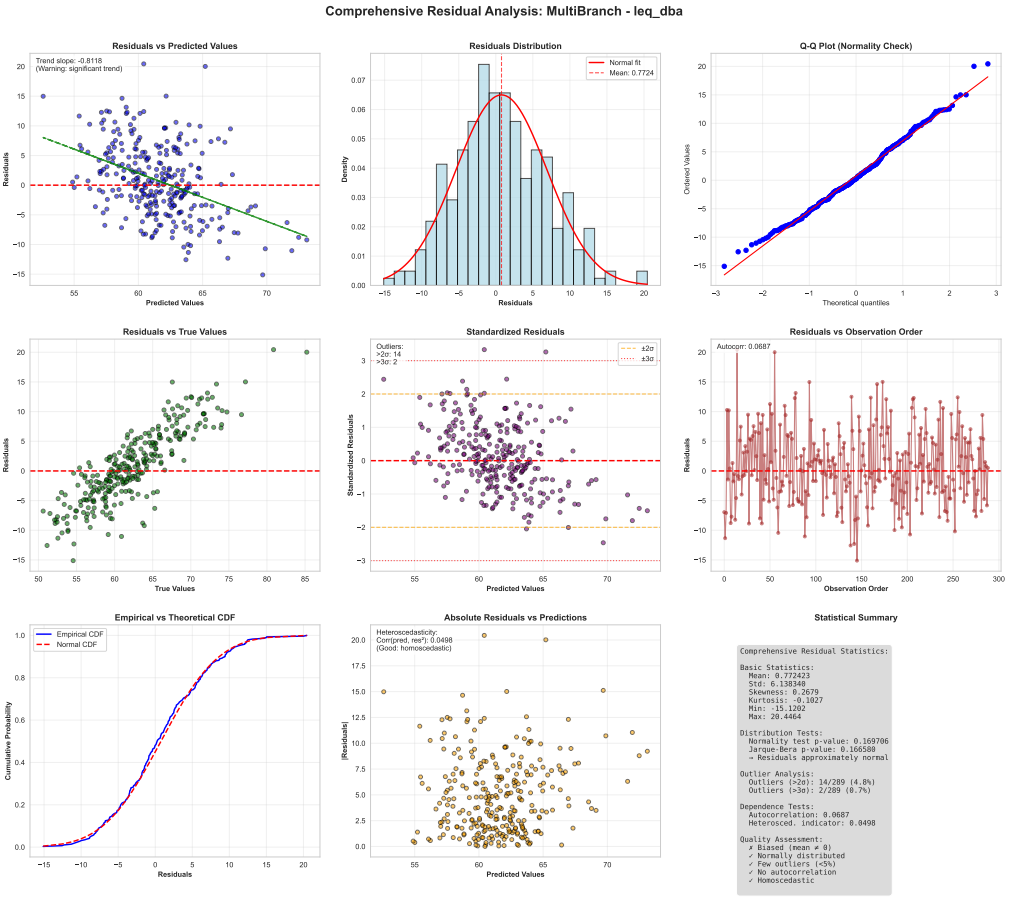


Figure A.64: Comprehensive Residual Analysis 7

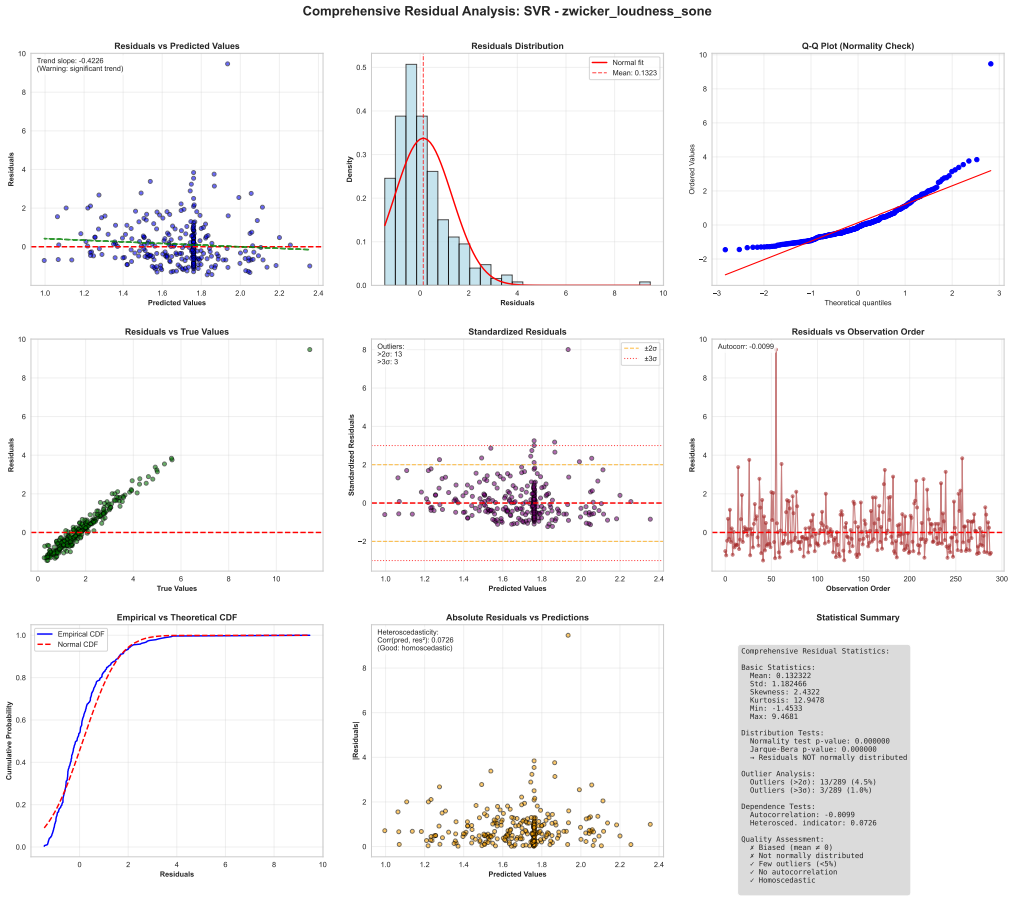


Figure A.65: Comprehensive Residual Analysis 8

A.6. Noise Map

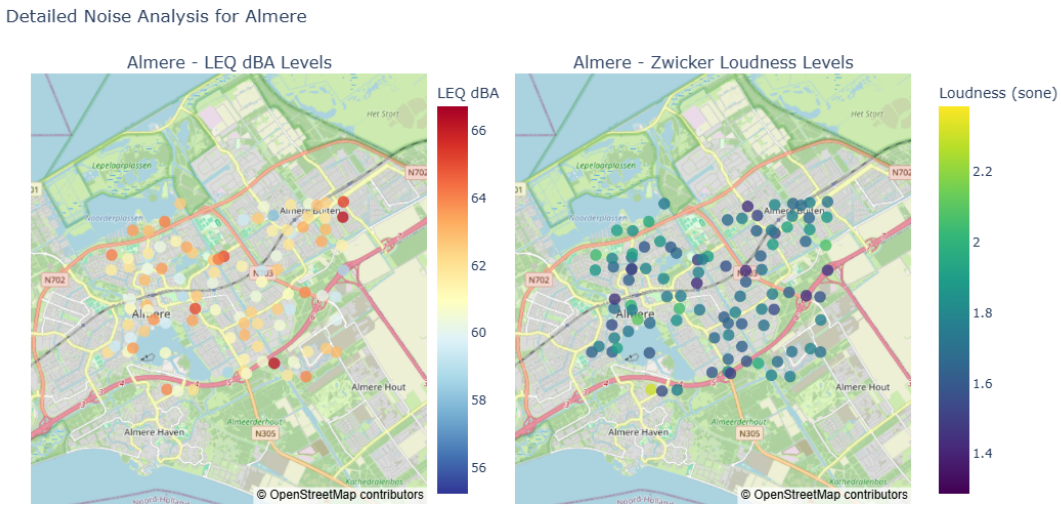


Figure A.66: Almere Detailed Noise Map

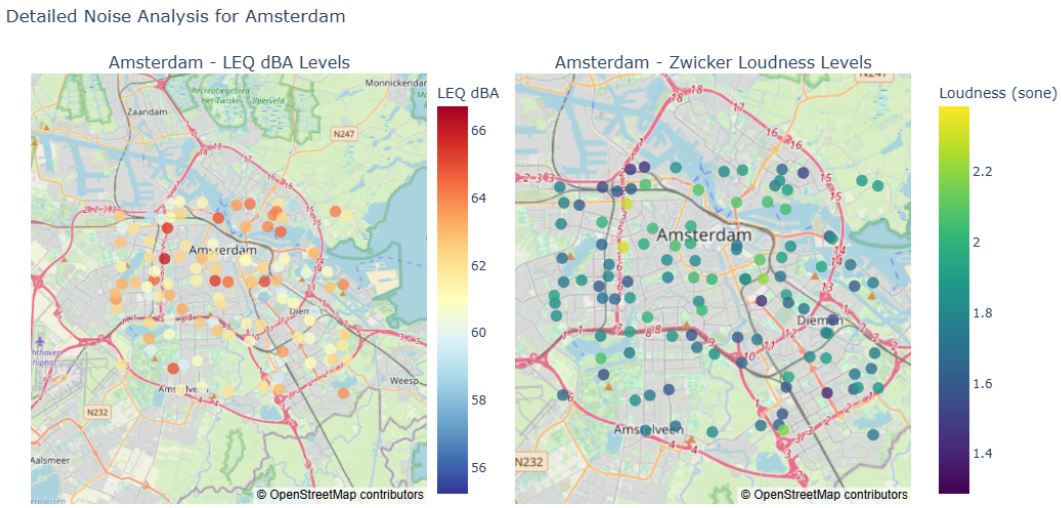


Figure A.67: Amsterdam Detailed Noise Map

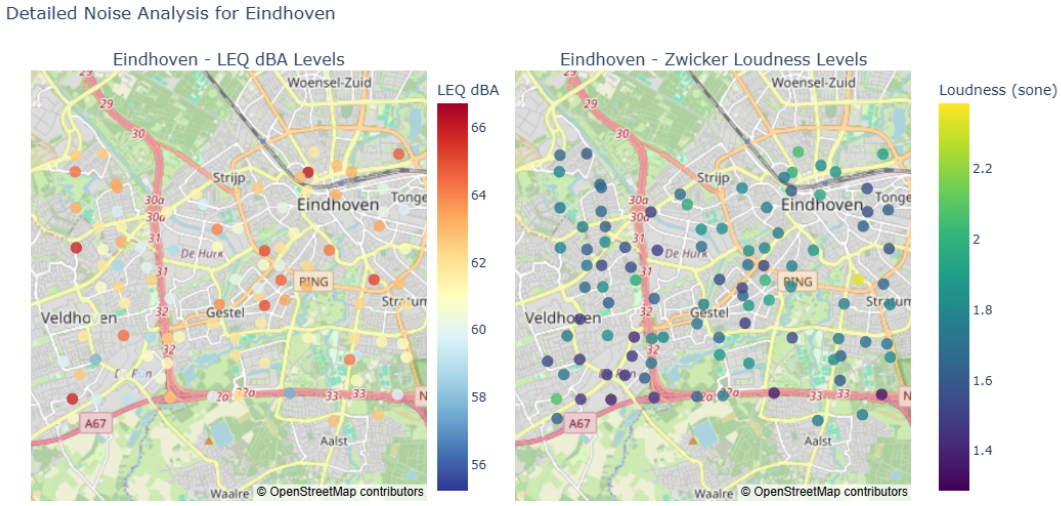


Figure A.68: Eindhoven Detailed Noise Map

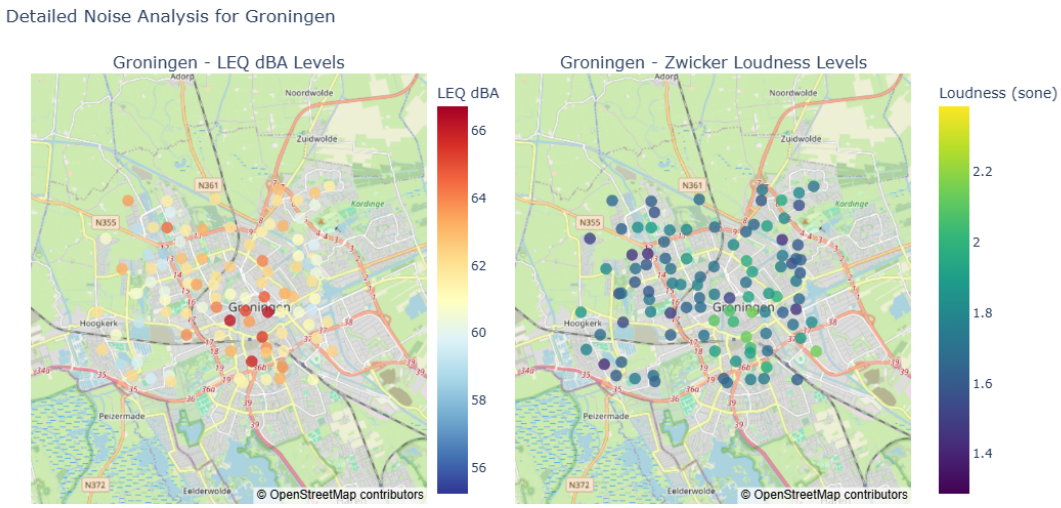


Figure A.69: Groningen Detailed Noise Map

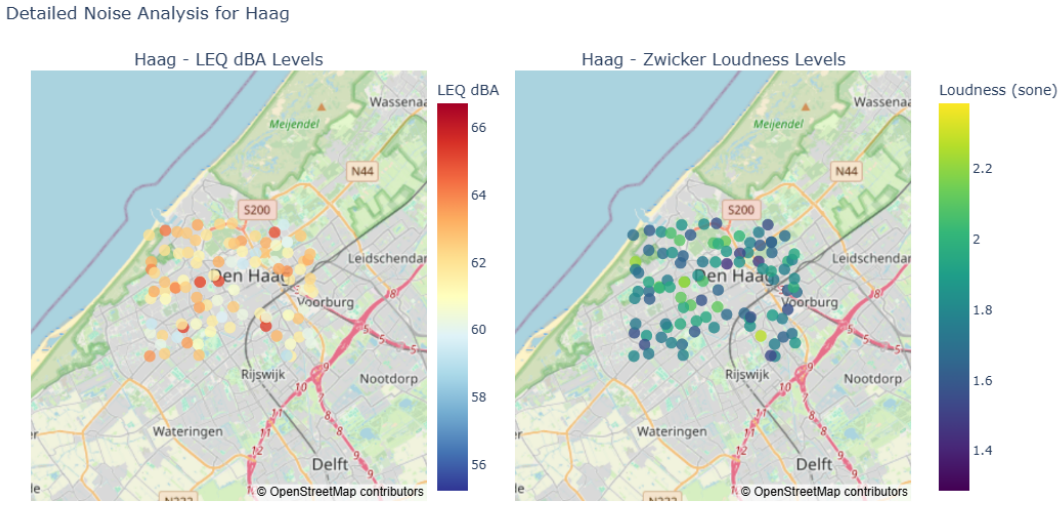


Figure A.70: Haag Detailed Noise Map

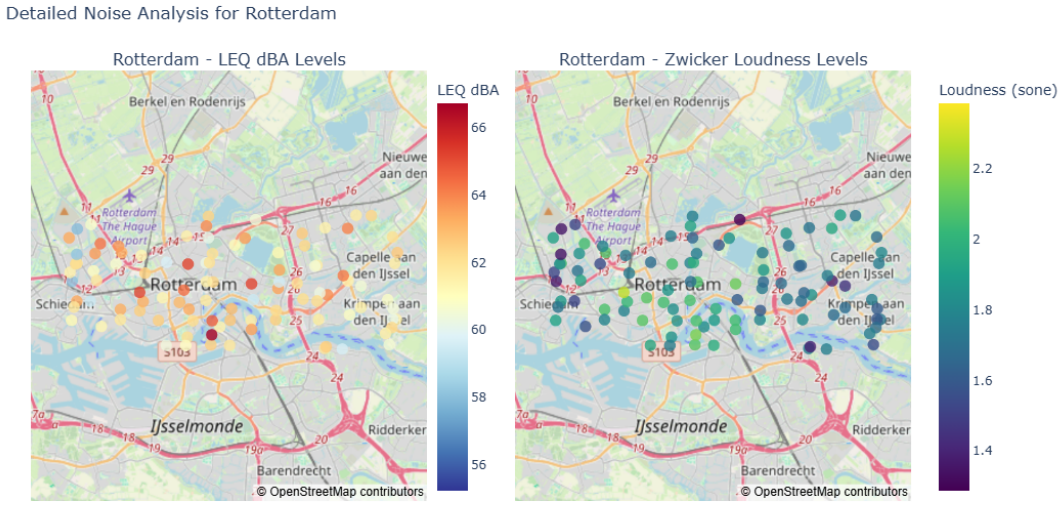


Figure A.71: Rotterdam Detailed Noise Map

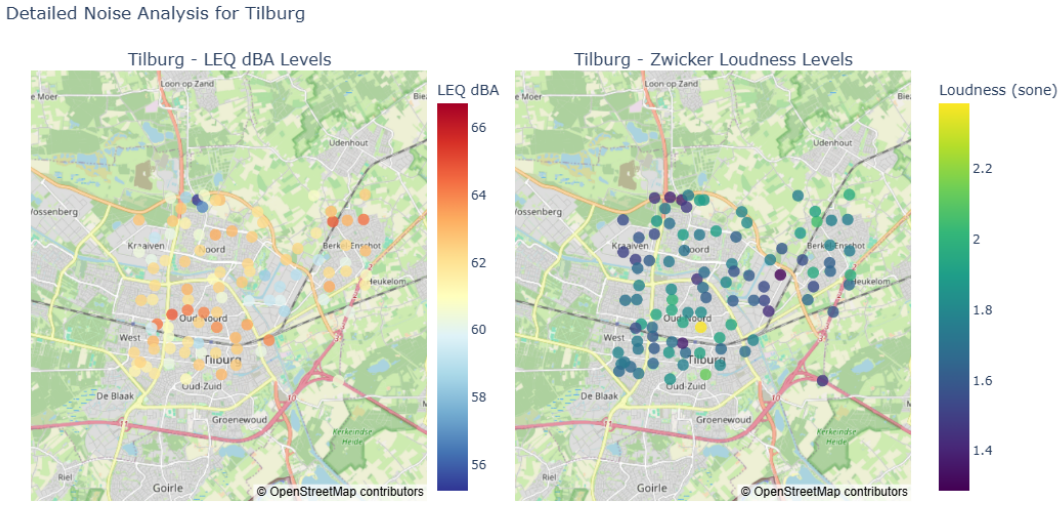


Figure A.72: Tilburg Detailed Noise Map

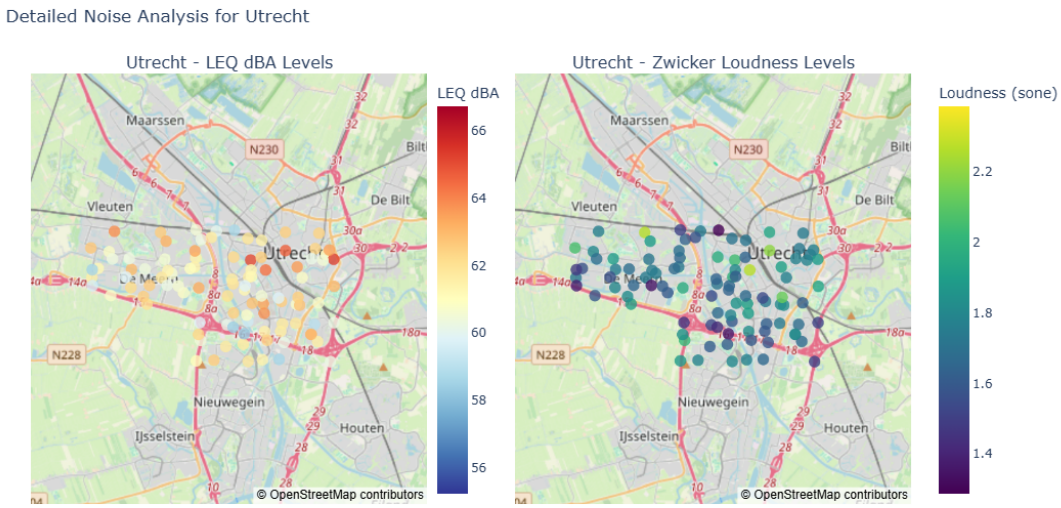
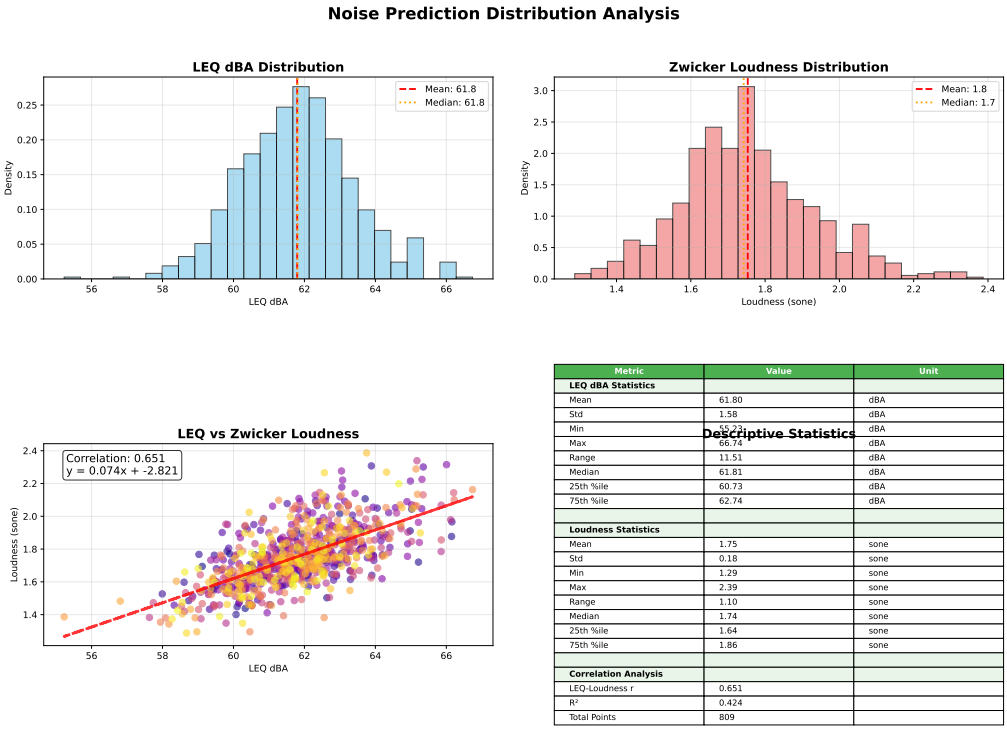


Figure A.73: Utrecht Detailed Noise Map



Metric	Value	Unit
LEQ dBA Statistics		
Mean	61.80	dBA
Std	1.58	dBA
Min	55.23	dBA
Max	66.74	dBA
Range	11.51	dBA
Median	61.81	dBA
25th %ile	60.73	dBA
75th %ile	62.74	dBA
Loudness Statistics		
Mean	1.75	sone
Std	0.18	sone
Min	1.29	sone
Max	2.39	sone
Range	1.10	sone
Median	1.74	sone
25th %ile	1.64	sone
75th %ile	1.86	sone
Correlation Analysis		
LEQ-Loudness r	0.651	
R^2	0.424	
Total Points	809	

Figure A.74: Noise Prediction Distribution Analysis

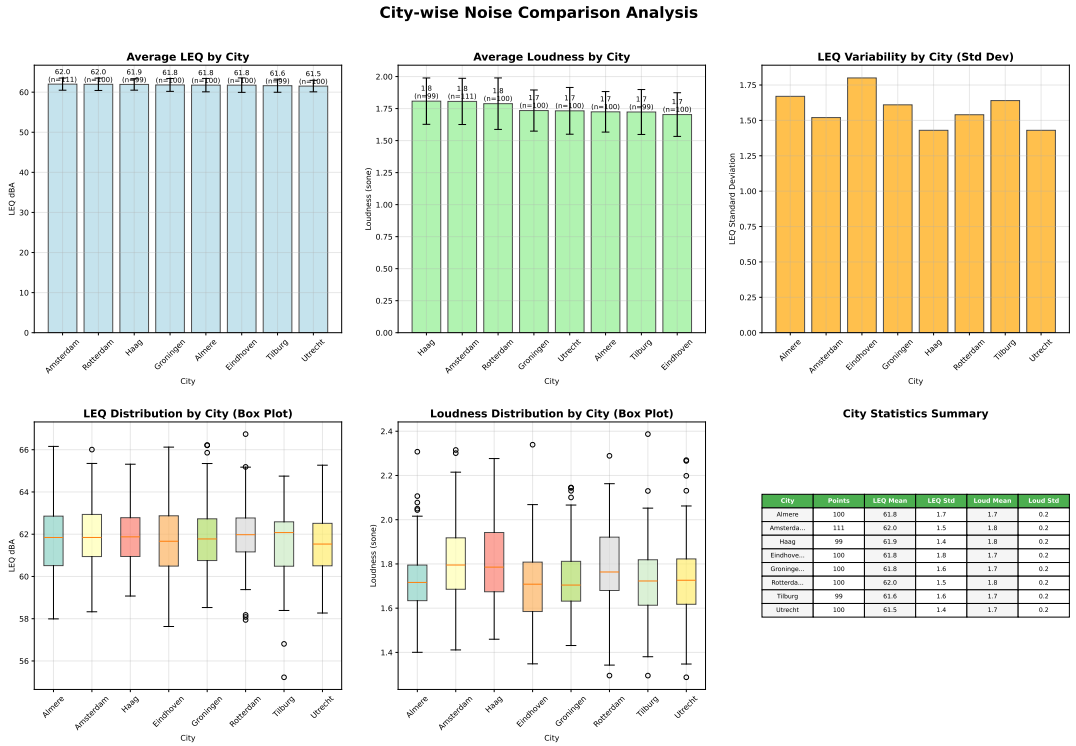


Figure A.75: Cross-City Average Noise Level Comparison Chart