Assessment of Machine Learning Models for the Prediction of Tire Rolling Noise Emission

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

M. Sc.

Michael Leupolz

aus Böblingen

Abstract

The tire road noise is the main traffic noise source in urban areas. For electric vehicles at speeds where aerodynamic disturbances do not result in large noise excitation, it even is the only relevant noise source. The health of a large percentage of the global population is impacted by traffic noise, which is why regulations for the vehicle noise are in place. Since the regulatory limits are stepwise decreased, vehicle manufacturers need to be able to estimate in the digital prototype phase if their vehicles meet the criteria. One large fraction thereby is the prediction of sound pressure levels caused by the tire on their specific test tracks. This dissertation investigates multiple machine learning approaches to predict sound pressure levels of rolling tires at one specific point of these measurement track, the whole measurement track, and as 1/3 octave levels for the entire measurement track. This is done on the basis of 1443 coast-by measurements, which include 42 different tires with tire nominal widths from 205 mm to 285 mm, rim diameters from 16 in to 22 in, and tire aspect ratios from 30 % to 65 %. All measurements were conducted at the Mercedes-Benz testing facilities in Immendingen, Germany. Predictions are done with linear regression models, lasso regression models, Artificial Neural Networks, Convolutional Neural Networks, a combination of Artificial Neural Network and Convolutional Neural Network, and Bayesian Hierarchical Models. It is shown, that with increased model complexity also the model accuracy improves without indications of overfitting. According to the model structure, different values are used as input features, models including convolutions receive tire tread pattern images extracted from 3D tire models reconstructed with the structured light approach. The shape of the contact patch is initially approximated as a rectangle. In later sections, contact patch shapes derived from the normal pressure distributions of individual tires are also used. Other models include features describing the tire, environment, and operation

conditions. The best leave-one-out model performance achieves a root mean squared error of 0.58 dB(A). The model used for this is the combined model, based on the contact patch shapes derived from the normal pressure distributions and the averaged sound pressure levels of both microphones. The model predicts the entire measurement distance and was minimized on the sum of the error of the test and training set. Through methods of explainable AI, it is shown that the most important tire features are the nominal width and the load index with positive and negative correlation to the sound pressure level. The load index also showed a strong correlation to the outer tire diameter.

Kurzfassung

Das Reifen-Fahrbahn-Geräusch ist die Hauptquelle für Verkehrslärm in städtischen Gebieten. Bei Elektrofahrzeugen ist es bei Geschwindigkeiten, bei denen aerodynamische Störungen nicht zu großen Geräuschanregungen führen, sogar die einzige relevante Lärmquelle. Die Gesundheit eines großen Teils der Weltbevölkerung wird durch Verkehrslärm beeinträchtigt, weshalb es Vorschriften für den Fahrzeuglärm gibt. Da die Grenzwerte schrittweise gesenkt werden, müssen die Fahrzeughersteller in der Lage sein, bereits in der digitalen Prototypenphase abzuschätzen, ob ihre Fahrzeuge die Kriterien erfüllen. Ein großer Teil davon ist die Vorhersage von Schalldruckpegeln, die durch die Reifen auf ihren spezifischen Teststrecken verursacht werden. In dieser Dissertation werden mehrere Ansätze des maschinellen Lernens untersucht, um Schalldruckpegel von rollenden Reifen für einen Punkt der Messstrecke, die gesamte Messstrecke und als 1/3 Oktavpegel für die gesamte Messstrecke vorherzusagen. Dies geschieht auf der Grundlage von 1443 Coast-by-Messungen, die 42 verschiedene Reifen mit einer Reifennennbreite von 205 mm bis 285 mm, einem Felgendurchmesser von 16 in bis 22 in und einem Reifenquerschnittsverhältnis von 30 % bis 65 % umfassen. Alle Messungen wurden auf dem Mercedes-Benz Testgelände in Immendingen, Deutschland, durchgeführt. Vorhersagen werden mit linearen Regressionsmodellen, Lasso-Regressionsmodellen, Künstlichen Neuronalen Netzen, Faltungsneuronalen Netzen, einer Kombination aus Künstlichen Neuronalen Netzen und Faltungsneuronalen Netzen sowie Bayes'schen Hierarchischen Modellen durchgeführt. Es wird gezeigt, dass sich mit zunehmender Modellkomplexität auch die Modellgenauigkeit verbessert, ohne dass es zu Overfitting kommt. Je nach Struktur des Modells werden verschiedene Werte als Eingangsmerkmale verwendet. Modelle, die Faltungen enthalten, erhalten Bilder von Reifenprofilen, die aus 3D-Reifenmodellen extrahiert wurden, welche mit

dem Ansatz des strukturierten Lichts rekonstruiert wurden. Die Form der Bodenaufstandsfläche wird zunächst als Rechteck approximiert, spätere Abschnitte enthalten auch Latschformen, die aus den Normaldruckverteilungen der einzelnen Reifen extrahiert wurden. Andere Modelle enthalten Merkmale, die den Reifen, die Umgebung und die Betriebsbedingungen beschreiben. Die beste Leave-One-Out-Modellperformance ergibt für die Wurzel des mittleren quadratischen Fehlers einen Wert von 0,58 dB(A). Das dafür verwendete Modell ist das kombinierte Modell, auf Basis der mittels Normalduckverteilung angepassten Bilder und dem gemittelten Schalldruckpegel beider Mikrophone. Das Modell sagt die gesamte Messstrecke vorher und wurde auf die Summe des Fehlers der Test- und Trainingsmenge minimiert. Mit Hilfe von Methoden der erklärbaren KI wird gezeigt, dass die wichtigsten Reifenmerkmale die Nennbreite und der Belastungsindex mit positiver und negativer Beziehung zum Schalldruckpegel sind. Der Belastungsindex zeigte auch eine starke Korrelation zu dem äußeren Reifendurchmesser.

Danksagung

Die vorliegende Arbeit entstand während meiner Anstellung als Doktorand im Bereich Außengeräusch bei Mercedes-Benz. Ich bedanke mich hiermit herzlich bei allen, die mich auf meinem Weg begleitet und unterstützt haben.

Vielen Dank Herr Prof. Dr. rer. nat. Frank Gauterin für die Betreuung dieser Arbeit und dem damit einhergehenden konstruktiven Austausch. Ich habe unsere Gespräche sehr geschätzt. Ebenso vielen Dank an Herrn Prof. Dr.-Ing. Wolfgang Kropp für die Übernahme des Korefferats.

Herzlichen Dank an Ralf Sperber, für das in mich gesetzte Vertrauen und die damit einhergehenden Freiräume die du mir geschaffen hast um dieses Promotionsprojekt erfolgreich zu bearbeiten. Vielen Dank auch an Michael Brandes für die Mercedes interne Betreuung meiner Promotion. An dieser Stelle auch allen Kollegen aus der Werkstatt, die mich bei meinen Messungen und den Messaufbauten unterstütz haben, herzlichen Dank.

Mein besonderer Dank gilt auch den Kollegen aus der Aerodynamik, Philipp Renz und Erich Jehle-Graf, welche mir großzügigerweise Rechenkapazität auf ihrem Cluster zur Verfügung gestellt haben und somit erst die für diese Dissertation nötigen Berechnungen ermöglicht haben.

Auch gesondert zu erwähnen ist Achim Winandi. Vielen Dank für deine Unterstützung als Forschungsgruppenleiter und die stete Bereitschaft auch eine spontane Rücksprache zu ermöglichen.

Vielen Dank Timo von Wysocki einerseits dafür, dass ich meine Masterarbeit bei dir schreiben durfte und du mich somit auch bei der Einstellung als Doktorand unterstützt hast. Andererseits für die danach noch überlappende

Promotionszeit und den auch über die Fertigstellung deiner Dissertation noch weiter bestehenden fachlichen und persönlichen Austausch.

Zu guter Letzt möchte ich meiner Familie und ganz besonders meinen Eltern für die Unterstützung und den Rückhalt über all die Jahre danken.

Michael Leupolz

Contents

Appendix

Notation

This chapter introduces the notation and symbols which are used in this thesis.

Symbols

a, b	Tire rolling noise coefficients
$\overline{a}$	Arithmetic mean of TRN coefficients a
$\overline{b}$	Arithmetic mean of TRN coefficients b
b_j	Bias of neuron j in a predefined layer
b_{lj}^n	Bias of neuron j in layer l for step n
b_{lj}^{n-1}	Bias of neuron j in layer l for step n-1
Δb_{lj}^{n-1}	Bias update of neuron j in layer l
C	Cost function
c_{ab}	Connector distribution of TRN coefficients initially sampled from $\mathcal{N}(42,5)$
c_{AR}	Tire aspect ratio
c_{GW}	Average circumferential groove width
c_{LI}	Tire load index
c_{BG}	Tire tread block-to-groove ratio
c_{RD}	Rim diameter
c_{TW}	Tire nominal width
d_1	Reconstructed instance
d_2	Reconstructed instance

$d_{1,i}$	i th sample for instance reconstruction
$d_{2,i}$	i th sample for instance reconstruction
D	Observed Data
f_{pitch}	Tire pitch frequency
f_0	Center frequency of 1/3 octave band
f_1	Lower frequency of 1/3 octave band
f_2	Upper frequency of 1/3 octave band
$f(x)$	Output of function f for input x
F	Set of all features
g	Gravitational acceleration
I_i	i th feature importance
K_{norm}	Regression Coefficient for temperature correction according to UNECE R117
K_{MB}	Regression Coefficient for temperature correction according MB correction
$l_{contact\ patch}$	Length of contact patch
L_{EAq}	A-weighted equivalent continuous Sound Pressure Level
L_i	Sound pressure level of measurement i
L_P	Sound pressure level
L_{pitch}	Tire pitch length
L_R	Reference sound pressure level of TRN
$\overline{L}$	Mean sound pressure level
m	Number of steps in Riemann Integral
$m_{vehicle}$	Vehicle mass
M_j	Model j
$\mathcal{N}(\mu_n, \sigma_n^2)$	Gaussian distribution with mean μ_n and variance σ_n^2
n_s	Number of samples
n_r	Number of measurements
n_F	Overall number of features

$\mathcal{O}$	Random perturbation of all features
$p_{\text{inflation}}$	Inflation pressure of tire
$P(\theta|D,M_j)$	Posterior Distribution
$P(D|\theta,M,_j)$	Likelihood
$P(D|M_j)$	Marginal Likelihood
$P(\theta|M_j)$	Prior
$P(D|M_j)$	Evidence
R^2	Coefficient of Determination
R_j^2	Coefficient of Determination for feature j
T_{air}	Air temperature
T_{road}	Road surface temperature
T_{ref}	Reference temperature according to UNECE R117
T_{tire}	Tire surface temperature
$\mathcal{U}(u_{\text{low}},u_{\text{high}})$	Uniform distribution with lower bound u_{low} and upper bound u_{high}
v	Vehicle speed in $\text{km}\,\text{h}^{-1}$
$\overline{v}$	Mean vehicle speed
v_i	Vehicle speed of measurement i
$v_{i,\text{log}}$	Logarithmic speed of measurement i with regards to v_{ref}
v_{min}	Min speed of data set
v_{max}	Max speed of data set
$v_{\text{m/s}}$	Vehicle speed in $\text{m}\,\text{s}^{-1}$
v_{ref}	Vehicle reference speed
$v_{\text{ref, norm}}$	Vehicle reference speed as defined in Regulation No. 117
w_i	Weight of neuron i
w_Z	Weight functions for possibility of entering coalition Z
W_{lij}^n	Weight matrix of neuron j in layer l for step n

W_{lij}^{n-1}	Weight matrix of neuron j in layer l for step n-1
ΔW_{lij}^{n-1}	Weight update matrix of neuron j in layer l
x_1, x_2, x_3	Cartesian coordinate system
x	Feature / Predictor variable
x	Input sample
X	Feature matrix / Predictor variable matrix
$\overline{x}$	Mean input sample
x'	Baseline of Integrated Gradients
x_i^{scaled}	Scaled values of x_i
y	Predicted value
y_i^{pred}	Predicted value for sample i
$\overline{y}^{\text{pred}}$	Mean predicted value for sample i
y_i^{ref}	Reference value for sample i
Y_{ji}	Predictive data of pooled model i for output j
Z	Coalition without feature j
$Z \cup \{j\}$	Coalition with feature j
α	Momentum
β	Regression coefficient
β_i	Regression coefficient of feature i
γ_1	Center point of average groove width parable
γ_2	Center point of block-to-groove ratio parable
ϵ	Residual
ϵ_i	Residual of sample i
ζ	Interpolation integrator
η	Learning rate
θ	Hyperparameters
θ_i	Hyperparameters of Model i
λ	Regularization parameter for Lasso regression
μ	Arithmetic mean

μ_π	Mean value for prior distribution
ν	Utility function /characteristic function
$\xi(\zeta)$	Linear path between two points
π	Prior
ϕ_j	Shapley value of feature j
$\phi_{j,i}$	i th sample for Shapley value of feature j
σ	Standard deviation
σ_a	Standard deviation of coefficient distribution of coefficient a
σ_b	Standard deviation of coefficient distribution of coefficient b
σ_P	Standard deviation of Sound Pressure Level (SPL) distribution
σ_x	Standard deviation of x
σ_π	Standard deviation for prior distribution
σ^2	Variance
τ	Temperature correction coefficient in Bayesian model
τ_{MB}	MB temperature correction coefficient

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
CB	Coast-by
CM	Combined Model
CNN	Convolutional Neural Network
CoD	Coefficient of Determination
CPB	Controlled Pass-by
CPX	Close-Proximity
DL	Deep Learning
FFT	Fast Fourier Transformation
IG	Integrated Gradients
KDE	Kernel Density Estimation
lasso	Least Absolute Shrinkage and Selection Operation
LeakyReLU	Leaky Rectified Linear Unit

LOOCV   Leave-One-Out Cross-Validation

MB   Mercedes-Benz

MBGD   Mini-Batch Gradient Descent

MCMC   Marcov Chain Monte Carlo

ML   Machine Learning

MSE   Mean Squared Error

OBSI   On-Board Sound Intensity

PDF   Probability Density Function

PSD   Power Spectral Density

ReLU   Rectified Linear Unit

RMSE   Root Mean Squared Error

SHAP   Shapley Additive Explanations

SPB   Statistical Pass-by

SPL   Sound Pressure Level

SRTT   Standard Reference Test Tire

STD   Standard Deviation

Tanh   Tangens Hyperbolicus

TRN   Tire Road Noise

TT   Test-Train

VIF   Variance Inflation Factor

WHO   World Health Organization

XAI   Explainable Artificial Intelligence

1 Introduction and Motivation

The exterior vehicle noise plays a crucial role in society, especially in urban areas. The European Environment Agency reported that 72 million inhabitants experience traffic noise levels of 55 dB in the day-evening-night period and 48 million people are exposed to nighttime levels of at least 50 dB in 2017 [Eur20]. However, the World Health Organization (WHO) strongly recommends 53 dB for the average road noise exposure and 45 dB for nighttime exposure. A connection between health issues as hypertension and ischaematic heart diseases as well as sleep disturbances was already recorded in 1994 [Hea94] and later reaffirmed [The11].

When it comes to engine and exhaust sound, some might consider these pleasant or enjoyable while others find them disturbing. In contrast to this, the Tire Road Noise (TRN) is widely considered as bothering. Yet many studies in recent years have shown that the overall exterior sound of a vehicle is dominated by tire road noise [Jab12, Put13, Zel18] and that the tire rolling noise already surpasses the propulsion noise at $35\,\text{km}\,\text{h}^{-1}$ to $50\,\text{km}\,\text{h}^{-1}$ depending on the vehicle [van05]. This domination solidifies itself through the electrification of vehicles, so that the main, if not only relevant noise source for future vehicles will be the TRN. Assuming that noises such as AVAS (acoustic vehicle alteration systems) are not required at higher Sound Pressure Levels (SPLs) and higher vehicle speeds. In highway regions, an additional noise due to aerodynamic effects can occur, this is however not expected to be as relevant as the TRN for urban regions where vehicle speeds are generally lower.

Even though, the health issues in relation with traffic noise and the domination of the traffic noise by TRN were often observed, no recent decrease in regulations specifically regarding the TRN are in place. In 2016 the necessary tire rolling noise was decreased between $2\,\text{dB(A)}$ to $5\,\text{dB(A)}$ for all passenger

vehicle tires. The decrease was defined in dependency of the nominal with, leading to a decrease of 4 dB(A) to 5 dB(A) in the range of 215 mm to 275 mm. [UNE22] These can be considered common tire specifications nowadays. But no later decreases than these specified in 2016 in regulatory noise levels for tires are known to the author.

For vehicle manufacturers, this is especially challenging since [UNE18] demands a decrease of pass-by SPLs of 4 dB(A) in three steps from July 2016 to July 2026. These decreases make it necessary for vehicle manufacturers to shift certification processes in the digital prototype phase to ensure passing the regulations. In general, this requires predictions of SPLs caused by engine, and exhaust for the entire measurement track as well as similar predictions for tire induced noise. However, as stated at the beginning of this section only the latter remains in case of electric vehicles which is why the focus of vehicle manufacturers lies on predicting the exterior TRN from digital prototypes. The fact that this is strongly driven by vehicle manufacturers, results from the stronger decrease in SPLs for vehicles and not for tires and from the observation that even test tracks with ISO 10844 surface vary in recorded SPLs [San17, San21]. This makes it necessary for vehicle manufacturers to fit regression models to their explicit certification tracks.

This dissertation explores various methods for predicting tire rolling noise levels using digital prototypes to facilitate digital certification processes. The methods used include standard linear regression models and Machine Learning (ML) models, which allow for the analysis of tire parameters' impact on SPL and the development of quieter tires. The primary objective is to design models with varying complexity to accurately simulate Coast-by (CB) measurements and determine the necessary complexity for faster tire selection during vehicle certification. Additionally, this dissertation aims to provide insight into the relevant parameters needed to predict the SPL.

2 Basics

This chapter is subdivided into five parts. First, influencing factors on the TRN are described based on previous publications. The focus herein lies on the impact of the tire and operational as well as environmental conditions. For analyzes of the road impact and acoustical excitation processes the reader is referred to [San80, San02, Li18a, Pin23]. Second, some statistical models for the TRN are presented. Third, mathematical fundamentals are explained, which is followed by ML models used in this research. The last part explains methods of Explainable Artificial Intelligence (XAI).

2.1 Literature Review of Tire Parameters and Operation Conditions on Tire Rolling Noise

One criterion in vehicle certification is the exterior noise. In vehicles with internal combustion engines generally three main sources of exterior noise are recorded: The engine, exhaust, and the tires. For electric vehicles the first two are mostly negligible so that the tire achieves a dominant role on the exterior vehicle noise [Jab12, Put13, Zel18]. However, literature indicates that even in vehicles with internal combustion engine the tire has a key impact [Put13].

The noise component of the tire is a result of the excitation through the contact with the road surface which leads to structural vibrations as well as direct disturbances in the air causing acoustical signals. The level of the recorded noise depends on many complex interactions. The following subsections provide an overview of different impact parameters subcategorized into tire parameters

and operating conditions. Following, the abbreviation TRN is used, which refers to the noise recorded for free rolling tires.

The following literature considers different measurement principles:

- Close-Proximity (CPX): Microphones are mounted in close proximity of the tire. The tire itself can be mounted in a trailer towed by a vehicle or on the vehicle itself. The microphones therefore move with the tire. [Deu17]

- Drum test bench: Microphones are mounted in close proximity of the tire. Tire rolls on test bench with different surfaces. [San02]

- Controlled Pass-by (CPB): Measurement setting as described in section 3.1. Measurement vehicle moves relative to microphones, which are mounted on the road. The vehicle drives at constant speed, the engine is not switched off. [San02]

- Coast-by (CB): Measurement setting as described in section 3.1. Measurement vehicle moves relative to microphones, which are mounted on the road. The principle is equivalent to CPB but the engine is switched off and gear switched to idle. [Int19]

- Statistical Pass-by (SPB): Measurement setting is similar to the one described in section 3.1. However measurements are conducted on standard roads under normal traffic conditions. Measurement vehicle moves relative to microphones, which are mounted on the road. [Deu22]

- On-Board Sound Intensity (OBSI): Two sound intensity probes are mounted in the close proximity of the tire. These move with the tire and have a predefined fixed distance to the tire. One is mounted approximately in the axis of the leading edge and one in the axis of the trailing edge. [Ame16]

2.1.1 Influence of Tire Parameters

The emitted sound is largely dependent on the tire itself. This is mainly due to the fact that different acoustic generation mechanisms act simultaneously, which can be amplified under different circumstances and tire parameters. In common literature the tire nominal width, aspect ratio, rubber hardness, tread pattern, rim diameter, tire diameter, and internal tire structure are investigated. Some literature on these is presented in the following paragraphs. Generally, it is observed that the separation of the above-stated influences is rather difficult since most of the parameters change interdependent. This complicates the separation of singular aspects. [San02]

The **tire nominal width** is of special interest since in most markets a trend of steady growth in width can be observed.

Sandberg et al. summarize literature from 1974 to 2001, and record an increase of $0.2\,\mathrm{dB(A)}\,(10\mathrm{mm})^{-1}$ to $0.8\,\mathrm{dB(A)}\,(10\mathrm{mm})^{-1}$ width increase in SPL. Hereby investigations on porous pavements show lower increase compared to dense pavements. [San02] A different study of over 400 tires on a drum test bench shows similar behavior on a smooth ISO surface. Tires are sorted into groups according to their outer diameter and the SPL is averaged for identical widths. The obtained averaged values increase by approximately $0.3\,\mathrm{dB}\,(10\mathrm{mm})^{-1}$. The same consistency could not be replicated on a rougher surface. [San02] Sandberg et al. put the overall expected increase in SPL to approximately $0.4\,\mathrm{dB}\,(10\mathrm{mm})^{-1}$ but state that for tires above $200\,\mathrm{mm}$ the absolute impact diminishes. [San02]

In 2011 research finds higher increases in SPL due to width increase for tires in the range of $145\,\mathrm{mm}$ to $205\,\mathrm{mm}$. 14 tires are categorized into three groups based on their outer diameter. Averaging the increase inside of the groups leads to $0.97\,\mathrm{dB}\,(10\mathrm{mm})^{-1}$ and $0.92\,\mathrm{dB}\,(10\mathrm{mm})^{-1}$ for a rough asphalt surface and an ISO 10844 surface for CB measurements. [Kum11]

These high values are supported by a finite element analysis of TRN conducted in 2013. The approach consists of two sub-models, a tire pavement model and a propagation model which are coupled by a fluid-structure interface. The

simulation is validated against measurements at $50\,\mathrm{km\,h^{-1}}$, $70\,\mathrm{km\,h^{-1}}$, and $90\,\mathrm{km\,h^{-1}}$. The tire model width is increased in increments of $10\,\mathrm{mm}$. This results in an increase in SPL of $1.0\,\mathrm{dB(A)\,(10mm)^{-1}}$ in the range of $180\,\mathrm{mm}$ to $200\,\mathrm{mm}$ and an increase of $0.2\,\mathrm{dB(A)}$ from $200\,\mathrm{mm}$ to $210\,\mathrm{mm}$. The authors assume that the width influence is negligible above $200\,\mathrm{mm}$. [Yan13] This generally coincides with the previous suggestion that for wider tires, the width effect declines. However, it is questionable if three steps in tire width in the range of $180\,\mathrm{mm}$ to $200\,\mathrm{mm}$ in the finite element model are sufficient to capture overall trends. Similarly, it could be argued that only one more step to $210\,\mathrm{mm}$ is not sufficient to truly verify that the increase in SPL through width increase gets smaller for wider tires. Furthermore, the incline of $1.0\,\mathrm{dB(A)}$ seems rather large compared to previously presented research.

The increase in SPL with tire nominal width is argued to result from three mechanisms. A stronger amplification through the horn effect, the necessity for more air displacement and therefore higher aeroacoustic induced noise, and a generally higher number of impacts of tread blocks during a fixed time period. The horn effect describes the acoustic amplification caused by the gradually increasing volume between the front or rear edge of the tire and the road surface. [San02]

In case of the **tire outer diameter** Sandberg et al. do not record a consistent correlation of SPL and diameter. This might be explained through opposing trends of air displacement and vibration induced mechanisms. Effectively, the angle between tire and pavement decreases for increased diameter. This might lead to a stronger horn effect. Simultaneously, a decreased angle also results in a more continuous deflection of the tread which is assumed to lower induced vibrations. [San02]

A decrease in SPL with increasing diameter of $0.065\,\mathrm{dB(A)\,mm^{-1}}$ is reported in an earlier investigation for tires in the range of $12\,\mathrm{in}$ to $13\,\mathrm{in}$. [Ejs82] However, it should be stated that tires of these diameters are rather outdated, so that it is questionable if the recorded trend also remains for greater diameters.

Regrouping the measurements according to the tire nominal width allows Kumar et al. to conduct an investigation of the diameter impact. The tendencies

for neither all groups nor both examined test tracks are consistent and the authors can only assume a slight negative correlation between tire diameter and SPL. [Kum11]

The **aspect ratio** puts the tire nominal width in relation to the tire sidewall height and is therefore an important characterization for each tire. However, no literature on the impact on the SPL is known to the author.

In regard to the **inner construction** of tires, one main aspect is the structure of the plies. However, since the 1980s radial ply tires dominate the markets [San02]. Analyzes comparing bias and radial ply are therefore no longer important.

The impact of altering the **belt stiffness**, on the other hand, is still of interest. First investigations on the belt stiffness are conducted by Ejsmont in 1982. The investigation includes eight tires on an asphalt road as well as on a drum test bench with a road replica and steel surface. Ranking tires across speeds ranging from $60\,\mathrm{km\,h^{-1}}$ to $120\,\mathrm{km\,h^{-1}}$, with steps of $20\,\mathrm{km\,h^{-1}}$ for various road surfaces, based on their A-weighted SPL, and subsequently averaging these rankings, reveals a negative correlation between belt stiffness and SPL. [Ejs82]

Tests of nine different tires with up to four different variations in inner construction show up to $1.5\,\mathrm{dB}$ difference on drum measurements on a road comparable surface at $80\,\mathrm{km\,h^{-1}}$ and $90\,\mathrm{km\,h^{-1}}$. [San02]

An effective yet simple way to gain some insight into the elastic properties of a vulcanized rubber is the investigation of the **shore hardness**. The scale ranges from zero indicating no stiffness to 100 representing a perfectly rigid material. [Cha07] Measurement of shore hardness are defined in DIN EN ISO 868 [Deu03].

A series of drum measurements on an ISO surface at $70\,\mathrm{km\,h^{-1}}$ show an increase in A-weighted SPLs with increasing shore hardness. The study includes three different tread patterns, one slick and two summer tires from which one is considered much less noisy. These tires are produced once with $57\,^\circ\mathrm{Sh}$ and

once with 67 °Sh. Since the recorded increase in A-weighted SPL is not consistent for each tread pattern, the authors assume that the increase positively correlates with the aggressiveness of the pattern. Aggressive patterns are defined as those with around 50 % block-to-groove ratio and many, wider grooves which are not circumferentially oriented. [San02] A study with more tires, selected from the open market, investigated the relation between shore hardness and SPL for three different drum surfaces. The tires are grouped into summer tires up to speed rating of $190 \, \text{km} \, \text{h}^{-1}$, summer tires between speed ratings of $190 \, \text{km} \, \text{h}^{-1}$ to $210 \, \text{km} \, \text{h}^{-1}$, summer tires with speed ratings above $210 \, \text{km} \, \text{h}^{-1}$, and winter tires. In all cases except the summer tires with speed ratings above $210 \, \text{km} \, \text{h}^{-1}$, the correlation between shore hardness and A-weighted SPL is positive, however the authors recommend leaving this group out due to its low variability in shore values and simultaneously high variance in other tire parameters. The correlation for different surfaces is similar. [San02]

During later research, the same authors reanalyze the data and put the impact of the shore hardness in the range of $1.0 \, \text{dB(A)} \, (10°\text{Sh})^{-1}$ to $1.5 \, \text{dB(A)} \, (10°\text{Sh})^{-1}$. The research also states that throughout the lifetime of a tire the shore hardness greatly increases. This is mainly due to high temperatures, which cause the rubber to deteriorate. The authors therefore point out a graphic from [Nor04] indicating that after 10 years the shore hardness for winter tires increases by 12 °Sh to 13 °Sh. [San07]

A more recent study of three different tires on three different surfaces performing CB measurements at $80 \, \text{km} \, \text{h}^{-1}$ also finds an increase of up to 3 dB(A) for changes from 50 °Sh to 70 °Sh. However, the author points out that values as low as 50 °Sh generally do not satisfy other criteria, such as handling and aquaplaning demanded by vehicle manufacturers for summer tires. [Sae08]

Bühlmann et al. investigate an approach to correct CPX measurements of Standard Reference Test Tires (SRTTs) defined by ISO/TS 11819-3:2017 (replaced by ISO/TS 11819-3:2021 [Int21]) due to changes in the shore hardness of tires. For this 247 shore hardness noise relations of previous literature are analyzed from which 172 are defined as statistically relevant. This leads to a median rubber hardness effect of $0.12 \, \text{dB} \, °\text{Sh}^{-1}$ for the SRTT defined for OBSI and CPX measurements. Temperature dependencies of the shore hardness are also

considered. [Büh18] This seems necessary since measurements in the range of 5 °C to 40 °C of five SRTTs show a decline in hardness of 0.25 °Sh °C^{-1} [Weh18]. In an earlier investigation Bühlmann et al. record an increase of 0.05 °Sh per measurement day for SRTTs [Büh13].

Lee et al. investigate the change in tire surface acceleration while driving over a cleat mounted on a drum test bench. They found that decreasing the hardness of the rubber as well as the tread stiffness leads to lower vibration energy. [Lee08] However these observations are in the range of 20 Hz to 100 Hz [Lee08] and therefore negligible for exterior noise.

The **tire tread pattern** seems to be the most obvious parameter that might be furthermore very easily adapted. Looking at the tire tread pattern many components should be considered.

The tire tread pattern is constructed out of many segments like blocks and grooves [San02]. The distance between two identical points on a block and its consecutive is called pitch and defines a length [Bri23]. Through variation of the pitch, the profile is **randomized** around the circumference of the tire. In close relation to the pitches, the offset is defined. It indicates an optimized circumferential shift of segments on both shoulder sides of the tire. [San02] The randomization of pitches is especially important for the subjective perception of noise since a constant pitch leads to constant impact frequency of blocks and therefore to a very narrow frequency band in the spectrum which indicates tonal character. In these cases the resulting frequency can be approximated through

$$f_{\text{pitch}} = \frac{v_{\text{m/s}}}{L_{\text{pitch}}}. \tag{2.1}$$

with f_{pitch} as frequency, $v_{\text{m/s}}$ as vehicle speed in m s^{-1} and L_{pitch} as pitch in meter. The randomization is said to be synchronous if the sequence of the pitches remains the same for both tire sides and asynchronous if the sequence changes. The asynchronous randomization generally allows for a more uniform mass and stiffness distribution. [San02]

Similar to the occurrence of aquaplaning, a tire tread pattern that is not properly ventilated will result in a much noisier tire. Essentially, the aim is to prevent air from being trapped between the rubber and the pavement, which leads to pressure fluctuations and thus to noise. This is achieved by designing the tread in such a way that all cavities are connected to the air surrounding the tire. Additional long grooves without ventilation should be avoided to reduce pipe resonance. Alternatively, Sandberg et al. mention a method developed by Bridgestone called *Groove Fence*, where a flexible rubber membrane is added to the grooves to break pipe resonance for an overall decrease of 1 dB(A) at 50 km h^{-1} in truck tires. [San02]

The impact of different tire tread patterns on the SPL was investigated in 1984 by Ejsmont et al. on a set of hand cut tires. Profiles are cut from nine 165 SR 13 slick tires, subsequently adding more characteristics after a measurement series on a steel drum and a road replica surface. One series includes changes of the transversal groove width from 2 mm, to 6 mm, to 9 mm, and to 12 mm. This shows an increase in the SPL up to 9 mm and a decrease in the SPL for groove width above 9 mm. A spectral analysis discovers an influence in the range of 1000 Hz to 4000 Hz. The highest observed changes are at around 1600 Hz independently of the vehicle speed. This increase could be explained by the pipe resonance frequency of the grooves. It is speculated that the decrease in SPL above 9 mm is caused by lower pipe resonance excitation due to less air pumping caused by better overall ventilation. This is supported by adding circumferential grooves which result in a substantial drop at the previous resonance frequency of the pipe resonance. As second parameter, the groove angles of 0°, 45°, 60°, 75°, and 90° are investigated. Circumferential grooves are represented by 0° and transversal grooves by 90°. The measurements show that changes from 90° to 75° lead to a decrease in the overall SPL while further decreasing the angle only results in a less profound tread impact frequency. Mirroring the diagonal grooves on the main plain of the tire does not result in any changes of the SPL. It is also demonstrated that adding a single circumferential groove for ventilation purposes to encapsulated grooves leads to a reduction between 5 dB to 8 dB in the SPL above 1000 Hz. In terms

of the groove orientation, the SPL decreases by 8 dB at the tread impact frequency if grooves are curved similarly to the trailing edge instead of the leading edge of the contact patch. Lastly, a decrease at frequencies above 4000 Hz for an increasing block-to-groove ratio is reported. An impact on the overall A-weighted SPL is not reported. [Ejs84] Considering the huge decrease in SPL due to the addition of a circumferential groove it should be mentioned that the reference tire in this case only had transversal grooves which are not ventilated by the shoulders and therefore build air pockets. These are generally considered noisy [San02].

As already stated by Ejsmont et al., these tendencies should not be taken for granted since the chosen patterns are rather generic [Ejs84] and, frankly do not represent realistic tire tread patterns, especially not for modern tires.

Furthermore, Schnieders analyzed the groove angle impact on the TRN for truck tires on a drum test facility. For this, grooves of the angles 0°, 15°, 30°, and 45° are carved in a semi slick truck tire. The author records a steady decrease in the SPL with increasing groove angle. He furthermore points out that the amount of change in the SPL resulting from adapted groove angles is also speed dependent. [Sch17] In contrast to Ejsmont et al. [Ejs84] this research defines the groove angles of 0° as transversal and 90° as circumferential groove.

In 2008 Saemann investigated the impact of circumferential and transversal voids. Two tires with identical tread patterns but different shore hardness are adapted in their circumferential and transversal void. For both tires, a decreased void volume from 20 % to 10 % in both directions leads to substantially lower recorded A-weighted SPLs for CB SPLs at 80 km h^{-1}. In case of the tire with 52 °Sh a decrease of 7.7 dB(A) and for the tire with 66 °Sh a decrease of 4.3 dB(A) is recorded if only the transversal void is decreased. The decrease from 20 % to 10 % circumferential void with constant 10 % transversal void for the softer tire leads to 2.5 dB(A) decrease in SPL. The author mentioned that this leads to a target conflict of aquaplaning and noise emission. [Sae08]

In the *tire road noise reference book* the authors also suggest that asymmetric tires could allow optimizing the outer side to generate lower noise while the

inward side is optimized for handling and braking characteristics. Especially in tires which also include a predefined direction of rotation, grooves can also be aligned in a manner to minimize sound excitation. [San02] However, this might not be an optimal solution since under sharp cornering maneuvers especially the outer shoulder of tires is under load and therefore needs to be optimized for handling and breaking.

Through various simulations conducted with the SPERoN software (see section 2.2), Liljegren concludes that the overall pattern stiffness should be kept constant and low to optimize noise emission. This is especially found in patterns with offset and S-shapes. However simulations are only conducted for $90 \, \mathrm{km \, h^{-1}}$. It is also found that offsets and randomization decrease the noise levels as well as higher groove angles to make voids more aligned with the tire's main plane. A further decrease in the A-weighted SPL is also achieved through designing transversal grooves thicker and shorter compared to narrow and long for constant voids. [Lil08]

In 2009, research includes additional resonators in the form of additional grooves in transversal grooves to actively dampen pipe resonances. An overall decrease for $80 \, \mathrm{km \, h^{-1}}$ of $0.8 \, \mathrm{dB(A)}$ is reported. [Fuj09]

More recent research linked the tire tread pattern noise recorded with OBSI to tire tread pattern spectra extracted from laser scans of tires. This proves that the recorded sound spectra after separating it from non-tread-pattern noise is actually defined by the tread pattern. [Li16]

Another research analyzed the impact of the groove resonance frequency on the A-weighted SPL by gluing acoustic foam in the grooves of a summer and a winter tire. Measurements for different loads, speeds, and different surfaces are conducted on a drum test bench. The delta between tire with and without foam averaged over all varied speeds, loads, surfaces, and tires is recorded as $1.7 \, \mathrm{dB(A)}$. The research also includes finite element models which are constructed from a deformable tire model and an imprint of the tire contact patch to simulate the tire groove resonance frequency and thus verify that the recorded delta in the SPL is in fact a result of the suppressed groove resonance. [Pin20]

Lastly an impact of tire wear and with it **tire profile depth** on noise emission is in place. Sandberg et al. summarize multiple studies that indicate an increase of SPL with increased wear and therefore lower profile depth. [San02] In contrast to that, a reduced SPL of 1 dB to 2 dB for a profile depth change from 8 mm to 2 mm through wear is recorded. [Ho13, cited from Sandberg U, Glaeser K. Effect of tire wear on noise emission and rolling resistance. In: Proceedings of inter-noise 2008. Shanghai, China, 26–29 October 2008.]

2.1.2 Influence of Operational and Environmental Conditions

The most important feature for the TRN is the **vehicle speed**. Generally, the correlation between speed and the SPL can be described as

$$L_p = a + b \log \frac{v}{v_{\text{ref}}} \tag{2.2}$$

where: a, b Tire specific coefficients in dB
        v Vehicle speed in km h^{-1}
        v_{ref} Reference speed in km h^{-1}

[on the basis of Nil76, San02]. Furthermore a strong linear relation between the coefficients a and b exists [San02]. The speed relationship had been proven through various literature for different surfaces as well as different measurement methods including the CB, CPX, drum, and SPB. This includes 1700 tire-surface combinations measured at the Technical University of Gdansk with the drum method and a total of four different surfaces. For these measurements a ranges from 4.3 dB to 58.7 dB with an average value of 30.6 dB and b ranges from 22.4 dB to 50.7 dB with an average value of 34.6 dB for an ISO surface when the vehicle speed v is normalized by v_{ref} as 1 km h^{-1}. [San02]

Another important factor for the TRN is the **tire load**. However, the effect of tire load is not completely consistent throughout literature.

Ejsmont reports an increase of 0.8 dB(A) to 1.5 dB(A) per doubling of load if the inflation pressure is not adjusted. An adjustment of the inflation pressure leads to an increase of 0.5 dB(A) to 1.0 dB(A) if the adjustments ensures doubling of the surface pressure. These values are reported for two tires (both 13 in, one with and one without tread pattern) measured on a drum test bench. The author also states that for more aggressive profile patterns, higher increases in SPL are possible in low pressure, high load scenarios. [Ejs82]

Sandberg et al. refer to an earlier investigation that found an increase of 2.4 dB(A) for a load increase from 1668 N to 3335 N. This results from measurements including more than 100 tires [San02, cited from Köllmann A., Ermittlung des Standes der Technik der Geräuschemissionen von PKW-Reifen, Research Report 105 05 144, FIGE GmbH, Herzogenrath, Germany, 1993]. Finally, Sandberg et al. conclude that in general the SPL increases with increasing load. However, the actual amount depends on the tire itself as well as other parameters as e.g. the vehicle speed. It is assumed that the increase in SPL does not exceed 2.5 dB(A) for passenger car tires in the range of 50 km h^{-1} to 80 km h^{-1} if the load is risen from 50 % to 100 % of the rated maximum load. [San02]

More recent research tries to investigate the SPL for different tire road pairings with a multilevel Bayesian regression. For this, measurements of three different tires are performed according to the CPX trailer method. The SPLs as well as vehicle speed, tire type, pavement type, and trailer weight are used as input for the ML model. This shows an increase in the SPL for increased load in three steps of 100 kg from 300 kg to 500 kg. The Bayesian model defines the mean impact as 0.026 dB(A) kg^{-1}. [Che21]

A previous publication of the author of this thesis investigated the load impact on CB measurements. For that, an identical tire is mounted on three different vehicles which differ 1020 N and 4650 N in weight. The vehicles are loaded with stone granulate to approximate axle loads of each of the other vehicles. During this process, no relevant change in the SPL due to vehicle or load adaptations for CB measurements in the range of 40 km h^{-1} to 90 km h^{-1} is recorded. [Leu22]

Another tire feature which is important for the acoustic emission of a tire is the **tire contact patch**. This includes the dimensions width, length, as well as shape and the thus resulting surface pressure. The literature presented here is not in each case connected to the TRN. However, since this dissertation also approximates the contact patch of different tires, the presented literature is within the scope of this research.

To analyze the impact of the contact patch, Polasik et al. increase the tire inflation pressure in steps of 0.5 bar between 0.5 bar and 3.0 bar and the tire load in steps of 50 kg between 50 kg and 500 kg. Surface measurements are performed for three different summer tires (165/80R14, 195/65R15, and 205/55R16). These show an overall trend of the contact area getting larger for higher load when the inflation pressure is kept constant. Also, the contact area decreases with increasing inflation pressure for constant tire load. Values at 2.0 bar especially show a linear relationship between tire load and contact area. The authors furthermore point out that at 2.0 bar and loads from 200 kg to 300 kg the contact area is unaffected by the tire itself. This might result from the correlating tire width and height. [Pol17]

Ekinci et al. confirm the trend of a growing contact area for larger loads and a smaller contact area for higher inflation pressure. The authors also show a linear relation between the contact area and the tire deflection and prove that the deflection is almost independent of the tire speed. [Eki15]
Sharma et al. furthermore mention the tire dimensions, carcass stiffness, and tread pattern design as important for the contact area. They also give an overview of analytical approaches to calculate the contact patches based on prior assumptions. [Sha96]

Riehm et al. present another method of estimating the contact area in 2019. This method is based on determining a deflection of the tire caused by load as a function of the footprint length and unloaded tire radius. The footprint length is approximated through a linear regression model, fitted with measurements. Combining the deflection equation with an equation representing the curvature over the tire width allows for calculation of a contact patch outline. [Rie19]

Other research investigates the effective contact area with a fast drying super emulsion ink on seven different tires of the size 155/70R12, five loads, and four inflation pressures. The effective contact area is the actual rubber surface that is in contact with the road. It therefore combines the outer silhouette of the contact area with the design of the tread pattern. From these measurements a regression model is built to calculate the effective contact area for an arbitrary load. The authors report a reduction of 17 % in the effective contact area for an increase in inflation pressure from approximately 1.4 bar to 2.1 bar and an area increase of 42 % for a load increase from 2 kN to 3 kN. The regression model shows that the data variance is explained to 51 % by the load, 37 % by the tire, and only 9 % by the inflation pressure. The authors assume the linear regression model to be sufficiently accurate to capture the relationship between effective contact area, tire load, and inflation pressure. [Mun14]
The reported impact of the tire on the contact area contradicts other presented literature stating that the tire itself has no noticeable influence on the contact area. However, in these publications it is not specified whether the effective contact area or the contact area is investigated.

Other research from Marshek et al. find similar values for the gross contact area for a mix of eight truck tires (bias ply and radial). They record a 30 % to 35 % increase in gross contact area for a 50 % increase in axle load and a decrease of 8 % to 20 % in gross contact area for a 50 % increase in inflation pressure. Changes in gross contact area for bias ply tires compared to radial tires caused by inflation pressure and load are similar, yet radial tires form overall smaller contact areas. [Mar85]

The size of the footprint can also impact the spectral representation of sound emission. This mainly results from the groove resonance effect in which standing waves occur in grooves which are open at one or both sides. The frequency of these waves is defined by the length of the grooves which in turn is influenced by the shape of the contact area. [Ulf04]

The impact of the **inflation pressure** on the contact area of the tire was already described. Other than the implicit impact on the emitted sound through the contact area, the inflation pressure was investigated in some literature on its direct impact on SPLs.

Iwao et al. report an increasing averaged SPL in dB(A) through applying torque from approximately $-150\,\text{N}\,\text{m}$ to $500\,\text{N}\,\text{m}$ on the tires. [Iwa96] With $380\,\text{N}\,\text{m}$ approximately equal to $1.5\,\text{m}\,\text{s}^{-2}$ and $620\,\text{N}\,\text{m}$ equal to $2.5\,\text{m}\,\text{s}^{-2}$, with the torque applied on each tire of a passenger car [San02]. The increase through torque is higher for an inflation pressure of 3 bar compared to 1 bar for a 195/60R14 tire under investigation. [Iwa96]

Another study on the effect of the inflation pressure on the recorded SPL does not show consistent behavior regarding its impact on the TRN. For this, three tires of size 135/70R13 are tested at the Technical University of Gdansk on two different surfaces mounted on a drum. Spectral investigations do not show clear correlations for specific frequencies. Only between $1250\,\text{Hz}$ to $2500\,\text{Hz}$ higher inflation pressures result in higher SPLs, however this behavior is only reported for one of three tires on an ISO surface. [San02]

The **temperature** has a strong impact on the recorded SPLs. This is not only explained by the change in shore hardness. Most literature separates three different temperatures: Air temperature, road temperature, and tire temperature [San02].

Research including ten different road surfaces for an overall of 200 SPB measurements found a positive correlation between the road and air temperature and constructed a linear regression model with a Coefficient of Determination (CoD) of 0.83. The measured road temperature deviates around $\pm 5\,°\text{C}$ from the estimated road temperature for a given air temperature. The measurements show that higher air temperatures lead to even higher road temperatures and vice versa. [And06]

Anfosso-Lédée et al. perform CPB measurements on a specific test track to investigate the impact of road, air, and tire temperature. The tire temperature is thereby the average of two measurements with a thermo-contact sensor in a groove and on a block. These measurements are performed before and after eight consecutive SPL measurements in the range of $70\,\text{km}\,\text{h}^{-1}$ to $110\,\text{km}\,\text{h}^{-1}$. Before starting the measurements the tires are heated up by driving for 15 min to 20 min. The investigated SPL is constructed similarly to eq. (2.7) but with a reference speed of $90\,\text{km}\,\text{h}^{-1}$. Two SRTTs are measured on seven different

road surfaces at four different days in between spring 2000 and winter 2001. The measurements show a linear relation between the temperatures

$$T_{\text{road}} = 1.7\, T_{\text{air}} - 4.5\,°\text{C}; \qquad T_{\text{tire}} = 1.05\, T_{\text{air}} - 15.8\,°\text{C} \tag{2.3}$$

and a CoD of 0.94 for both regressions if all road types are included. The relation between the SPL and the air temperature is calculated to $-0.10\,\text{dB(A)}\,°\text{C}^{-1}$ for dense bituminous and $-0.06\,\text{dB(A)}\,°\text{C}^{-1}$ for porous bituminous pavements at $90\,\text{km}\,\text{h}^{-1}$. The research also discovers differences in the temperature dependencies of different 1/3 octave bands between $100\,\text{Hz}$ to $5000\,\text{Hz}$. Temperature dependency merely exists below $500\,\text{Hz}$ and above $1600\,\text{Hz}$. The authors believe that the temperature importance at higher frequencies results from changes in adhesion. Altogether, these temperature dependencies indicate that it is more plausible to speculate that the temperature changes impact the importance of different generation mechanisms and not the propagation mechanisms. [Anf07]

Bendtsen et al. states that the temperature effect on the SPL highly depends on the tire itself as well as the pavement type. In case of OBSI measurements on SRTT Bendtsen et al. report $-0.027\,\text{dB}\,°\text{C}^{-1}$ for air temperature correction and $-0.018\,\text{dB}\,°\text{C}^{-1}$ for pavement temperature correction averaged for two different pavement types. They also speculate that the effect of different pavements cancels out if enough tires are included in the measurements. [Ben09]

Bühlmann et al. investigate the temperature impact during CPX measurements. Two tires (truck and SRTT) are tested on 39 road surfaces at the speeds $50\,\text{km}\,\text{h}^{-1}$ and $80\,\text{km}\,\text{h}^{-1}$. Linear relations between the overall SPL and different temperatures are analyzed. The average CoD is between 0.86 to 0.87 for air temperature measured at $15\,\text{cm}$, $35\,\text{cm}$, $150\,\text{cm}$ above the surface, surface temperature, and tire temperature. The linear correlations of the different temperatures with the SPL are similar. Also all temperatures show high linear correlation with the air temperature measured at $150\,\text{cm}$. This leads the authors to recommend using the air temperature measured at $150\,\text{cm}$ above the surface for correction since it is the least sensitive to solar radiation. In contrast to other presented research, a temperature effect was also found in

the mid frequency range from 800 Hz to 1250 Hz. As overall average correction slopes of the SRTT, $-0.10\,\text{dB(A)}\,°\text{C}^{-1}$ for dense, $-0.08\,\text{dB(A)}\,°\text{C}^{-1}$ for concrete, and $-0.05\,\text{dB(A)}\,°\text{C}^{-1}$ for porous pavements are recorded. [Büh11]

In a review which combines 17 earlier published studies, Bühlmann et al. propose different temperature correction parameters for different pavement categories. All studies are selected due to their standard error of the temperature slope and CoD of the regression. Studies that reported only the road temperature are converted to air temperature through conversion factors. The 17 reviewed studies include CPX, OBSI, CPB, SPB, and L_{EAq} measurements, speeds from $50\,\text{km}\,\text{h}^{-1}$ to $140\,\text{km}\,\text{h}^{-1}$, car and truck tires, with the majority being car tires. The review concludes grouping the pavements into dense, cement concrete, and porous pavements with $-0.10\,\text{dB(A)}\,°\text{C}^{-1}$, $-0.07\,\text{dB(A)}\,°\text{C}^{-1}$, and $-0.05\,\text{dB(A)}\,°\text{C}^{-1}$ for temperature slopes. [Büh14]

It can be concluded, that the overall observation in literature is to correct the SPL according to the temperature with higher temperature leading to lower SPL.

This probably leads to a temperature correction allowed during tire measurements according to UNECE regulation no. 117 [UNE22] with

$$L_R(T_{\text{ref}}) = L_R(T_{\text{road}}) + K_{\text{Norm}}(T_{\text{ref}} - T_{\text{road}}) \tag{2.4}$$

where: T_{ref} Reference temperature of 20 °C
  T_{road} Road surface temperature in °C
  K_{norm} Regression coefficient
  L_R TRN reference sound level in dB(A)

and

$$K_{\text{norm}} = \begin{cases} -0.03\,\text{dB(A)}\,°\text{C}^{-1} & \text{if } T_{\text{road}} > T_{\text{ref}} \\ -0.06\,\text{dB(A)}\,°\text{C}^{-1} & \text{if } T_{\text{road}} < T_{\text{ref}} \end{cases} \tag{2.5}$$

The L_R value is the calculated reference sound for multiple measurements calculated trough

$$L_R = \overline{L} - \beta_R \overline{v} \tag{2.6}$$

$$\overline{L} = \frac{1}{n_R} \sum_{i=1}^{n_R} L_i \tag{2.7}$$

$$v_{i,\,\log} = \log \frac{v_i}{v_{\text{ref, norm}}} \tag{2.8}$$

$$\overline{v} = \frac{1}{n_R} \sum_{i=1}^{n_R} v_{i,\,\log} \tag{2.9}$$

$$\beta_R = \frac{\sum_{i=1}^{n_R} (v_{i,\,\log} - \overline{v})(L_i - \overline{L})}{\sum_{i=1}^{n_R} (v_{i,\,\log} - \overline{v})^2} \tag{2.10}$$

where:
L_i	SPL of measurement i
L_R	TRN reference sound level in dB(A)
$\overline{L}$	Mean of L_i of n_R measurements
v_i	Speed of measurement i
$v_{i,\,\log}$	Logarithmic speed of measurement i with regards to v_{ref}
$v_{\text{ref, norm}}$	Reference speed used in Regulation No. 117
$\overline{v}$	Mean of $v_{i,\,\log}$ of n_R measurements

and $v_{\text{ref, norm}}$ as $80 \, \text{km h}^{-1}$. [UNE22]

However, in a previous study it could be found that a logarithmic approach might fit the temperature gradient better than the unsteady slope adaptation of two straight lines [Leu22]. A recent *Working Party on noise and Tyres* [UNE23b] also considers changing the temperature correction to a logarithmic curve [UNE23a].

A recent investigation considers the temperature impact for urban measurements. The L_{EAq} according to IEC 61672-1 [Int13] is calculated for a defined time period in Madrid traffic in which the traffic flow could be considered stable. The microphones are placed at 4 m to 6 m above ground in the perimeter of traffic lanes. For 21 different measurement locations, ranges from $-0.04 \, \text{dB} \, °\text{C}^{-1}$ to $-0.13 \, \text{dB} \, °\text{C}^{-1}$ with CoD ranging from 0.04 to 0.42

are found. The L_{EAq} are calculated for hourly intervals, with 1236 to 1539 available intervals for different measurement locations. [Bar22]

2.2 Literature Review Tire Noise Statistic Models

Due to the high complexity of the TRN, many approaches to estimate the TRN are considered in research. These include finite element simulations and statistical models. Some models also actively take a step back from the total SPL and focus on frequency levels which allows for an investigation of single acoustical phenomena. Some of these approaches are presented below.

Kim et al. investigate an approach to predict the air-pumping noise for rolling tires. They model the acoustical source by what they call a piston/sliding-door/cavity geometry. In essence, this represents a transversal groove which is continuously closed and opened by sliding a flat surface over it while compressing and expanding the air inside the groove through moving its bottom. The flow properties are obtained from a Navier-Stokes finite volume simulation for the piston cavity. These properties are propagated to a larger CFD simulation to compute the flow field in the near field of the tire. Through a Kirchhoff integral, far field SPLs are simulated. [Kim06]

A hybrid approach, which combines statistics and physical models is the SPERoN (Statistical Physical Explanation of Rolling Noise) model. The model calculates 1/3 octave levels for CB measurements. As input, road and tire properties as well as a 3D geometry are necessary. The calculation relies on splitting the overall sound excitation into four subparts: Mechanical induced vibration, airflow within the contact patch, interior cavity noise of the tire, and aerodynamic noise induced by the air flow around the car body. Thus assuming incoherent processes, allowing linear superposition of intensities. Each of the subparts is modeled as physical equation with regression coefficients. The regression coefficients are estimated on the basis of the Sperenberg project data. [Bec08] The model is validated for one tire and eight surfaces, the difference in overall A-weighted SPL is recorded as

1 dB(A) on average and 2.5 dB(A) as maximum value [Kle08]. A deviation over the measured distance is not recorded. The focus of SPERoN lies on optimizing the road surface [Bec08]. The model allows for predictions in the range of 315 Hz to 2000 Hz and from 50 km h^{-1} to 120 km h^{-1}. To calculate the vibration excitation of the tire, the model relies on calculating its contact forces through a plane of uncoupled springs (Winkler bedding) and an orthotropic plate as coupled tire module. These demand tire properties that can be obtained from vibroacoustical measurements. [Bec12] Forssén et al. use the SPERoN model as basis for an auralization model of pass-by simulations. The method combines it with an adapted version of the earlier developed auralization approach for pass-by measurements which is a result of the *LISTEN: Auralization of Urban Soundscapes* project. [For18]

In 2017 Li et al. predict the tread pattern related noise with an Artificial Neural Network (ANN). For this, 23 tires with similar size and aspect ratio are measured with the OBSI method. The ANN receives two different types of spectra as input for each tire. A coherent tread pattern spectrum and an air volume velocity spectrum of the contact patch. These are both extracted from 3D tire tread patterns obtained through laser scanning the tires. The coherent tread pattern spectrum is obtained through averaging the complex Fourier coefficients of each circumferential line on the tire, before calculating the power spectrum. The air volume velocity spectrum is obtained by moving a rectangle representing the contact patch over the 3D tire. From the 3D model, the air volume in the grooves for each position of the rectangle can be calculated. Deriving this signal by the time leads to a volume velocity signal which can be Fourier transformed to calculate the air volume velocity spectrum. The maximum recorded difference in the tread pattern noise is recorded as 17.2 dB(A). The authors separate the entire SPL into two parts, the noise related to the tread pattern and the non-tread pattern related noise through an order analysis. The tread pattern related noise is described through the coherent tread pattern spectrum and the air velocity spectrum. The ANN reaches an average error of 2.1 dB and 3.5 dB for test and validation sets. [Li17]

Other research investigates an image based approach to predict the air pumping noise. The essence of the algorithm is to track air gaps that enter and exit

the contact patch. This allows for a calculation of the change in air volume in the contact patch. In combination with an assumed tread deformation due to the load, this allows the modeling of a monopole sound source under free-field conditions. The algorithm is then used to investigate the impact of transversal groove length, width, angle, contact patch length, offset, and vehicle speed on the SPL. [Oor17]

Chen et al. design a multilevel Bayesian analysis on the basis of trailer measurements according to the CPX norm. During the measurements, the trailer weight, vehicle speed, tires, and road surface are changed. These include three different tires and 38 different road sections which can be subdivided in five categories. The multilevel Bayesian model includes parameters to describe the road surface in more detail, in case of the tires this is not the case. These measurements and the model are focused on analyzing impact factors on the noise contribution with main investigations in the area of different pavements. [Che21].

Lee et al. predict the TRN with an ANN and a Convolutional Neural Network (CNN). The networks are based on drum test bench measurements of 28 tires at $80 \, \text{km h}^{-1}$ with three microphones at a radius of 1 m from the center of the tire. The microphones are all located in the plane spanned by the direction of travel and the tire's axis of rotation. Four of the 28 tires have the dimension 215/60R16, 24 have the dimension 245/45R18, yet they differ greatly in their tread patterns. The ANN receives the coherent tread profile spectrum as well as the contact patch air volume velocity spectrum after sampling both down to order spectra and applying Gaussian curve fitting. The spectra are calculated on the basis of laser scans of the tires, the ANN predicts the tread pattern noise spectra for 101 orders. In case of the CNN tread pattern images are compressed with 2D wavelet transformation. These are then passed through the CNN to predict the same 101 orders as in the ANN investigation. Four specific tires are left out of the training set for testing. For model evaluation CoDs between target and predicted values of 0.745 for the ANN and 0.89 for the CNN are reported. [Lee21]
However, no test error in terms of SPL as e.g. Root Mean Squared Error (RMSE) is reported.

A more detailed research on different ML algorithms to predict the TRN is presented in 2022. CB measurements are performed in the range of $70\,\mathrm{km\,h^{-1}}$ to $90\,\mathrm{km\,h^{-1}}$ to fit these to the tire noise label calculated for $80\,\mathrm{km\,h^{-1}}$ for 23 tires. The authors investigate four different model types: support vector machine, relevance vector machine, ANN, and CNN. All models but the CNN receive the tire load index, tire speed index, tire nominal width, aspect ratio, rim diameter, vehicle speed, frequency corresponding to the SPL to predict, tread impact spectra, and air pumping spectra. In case of the CNN the last two are replaced by a tread pattern image and the groove depths. The output of the models is the SPL at the frequency used as input. Through this, the range of $250\,\mathrm{Hz}$ to $2500\,\mathrm{Hz}$ is covered. The performance of each model is estimated with k-fold cross-validation leading to RMSE values of the total SPL of $1.21\,\mathrm{dB(A)}$, $0.83\,\mathrm{dB(A)}$, $1.48\,\mathrm{dB(A)}$, and $1.19\,\mathrm{dB(A)}$ for the test losses of above-mentioned models. [Moh22]

Rapino et al. predict the 1/3 octave sound intensity levels by training an ANN on data of 83 tires measured at a drum test bench with sandpaper in a semi-anechoic chamber. The measurements are performed at the speeds of $50\,\mathrm{km\,h^{-1}}$, $80\,\mathrm{km\,h^{-1}}$, and $110\,\mathrm{km\,h^{-1}}$ and SPLs are recorded by six microphones placed in a semicircle with radius $1.5\,\mathrm{m}$ around the tire. The microphones are all located in the plane spanned by the direction of travel and the tire's axis of rotation. 24 parameters are used as input for the network. These are the operation parameters: RPM of the tire and inflation pressure; Footprint parameters: Width, length, and roundness of the footprint; Tire structure parameters: Width, external radius, and rubber hardness; Tread pattern characteristics: 1/3 octave levels in the range of $200\,\mathrm{Hz}$ to $4000\,\mathrm{Hz}$. [Rap23] The last parameter is similar to the process described by [Li17]. Three different speeds are used as training samples for each tire. This leads to 249 data points. The ANN structure can be considered small with only one hidden layer with fourteen neurons. A sigmoid is uses as activation function. The data are split in training, validation, and test data set by 70 %, 15 %, and 15 % respectively. The validation data are used for stopping the training early to decrease overfitting. The test loss is recorded as $2.3\,\mathrm{dB(A)}$, the output values are fourteen 1/3 octave bands sound intensity values in the frequency range of $200\,\mathrm{Hz}$ to $4000\,\mathrm{Hz}$. [Rap23]

A more comprehensive review which also considers earlier research on TRN models can be found in [Li18b].

2.3 Statistical Parameters

This section provides an introduction to statistical parameters used in the presented research.

Root Mean Squared Error (RMSE)

The RMSE is often used as evaluation criteria for model performance. For discrete functions it can be calculated through

$$RMSE = \sqrt{\frac{1}{n_s} \sum_{i=0}^{n_s} \left(y_i^{\mathrm{pred}} - y_i^{\mathrm{ref}} \right)^2} \qquad (2.11)$$

where: y_i^{pred} Predicted value for sample i
            y_i^{ref} Reference value for sample i
            n_s Number of samples

In models with multiple outputs, this can also include vectors. In these cases often the mean of the RMSE of the output variables is reported as RMSE. In the present dissertation, the RMSE is always calculated for vectors without averaging over the amount of vector dimensions.

Coefficient of Determination (CoD)

Another method to determine how well a model represents the data is the CoD also referred to as R^2. The value space ranges from 0, indicating no relation, to 1, indicating perfect data fit. It is calculated through

$$R^2 = 1 - \frac{\sum_i^{n_s} \left(y_i^{\mathrm{ref}} - y_i^{\mathrm{pred}} \right)^2}{\sum_i^{n_s} \left(y_i^{\mathrm{ref}} - \overline{y}^{\mathrm{pred}} \right)^2} = 1 - \frac{\sum_i^{n_s} \epsilon_i^2}{\sum_i^{n_s} \left(y_i^{\mathrm{ref}} - \overline{y}^{\mathrm{pred}} \right)^2} \tag{2.12}$$

where:
R^2		Coefficient of Determination
y_i^{pred}		Predicted value for sample i
y_i^{ref}		Reference value for sample i
$\overline{y}^{\mathrm{pred}}$		Mean output value for all samples
n_s		Number of samples
ϵ		Residual

More graphically, the CoD can be interpreted as the proportion of the variance in output data that is explained by the model. However, models should not be chosen solely on their CoD. This is due to the fact, that $\lim_{j \to n_s} R^2 = 1$ with j representing the number of features. Which indicated that the CoD increases through ever-growing model complexity.

Variance and Standard Deviation

The variance is defined through

$$\sigma^2 = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (x_i - \mu)^2 \tag{2.13}$$

where:
σ^2	Variance
n_s	Number of samples
x_i	Data point
μ	Arithmetic mean

The Standard Deviation (STD) is defined as σ. It indicates how much the data differs on average from the mean value and therefore is an important measurement for the overall data spread [Boy18].

Variance Inflation Factor

The Variance Inflation Factor (VIF) is a statistic parameter to estimate if multicollinearity of input features is in place. It is calculated through

$$VIF_j = \frac{1}{1 - R_j^2} \tag{2.14}$$

with R_j^2 as the CoD that is obtained for feature j predicted with all remaining features. A value of 1 for the VIF indicates no correlation. [Mon12]
Even though the VIF gives an intuitive understanding of the proportions of variance in the data, the threshold which is considered to indicate severe multicollinearity is not clearly defined. O'brien combines literature which mostly indicate that the VIF should not exceed 10, however some literature refers to 4 as critical. Altogether, O'brien makes a case that the VIF should always be considered in the context of the specific problem since factors such as e.g., the amount of samples also impact the variance and with it the VIF. [Obr07]
Others identify values between 5 and 10 as possibly problematic and values above 10 as problematic [Aki15].

2.4   Machine Learning

Machine Learning (ML) is a subset of methods belonging to the broader spectrum of Artificial Intelligence (AI). It can be thought of as computational methods that learn from previous data to improve their accuracy or performance. The use case of these methods has almost no limit. [Moh18] One common problem in ML is the model selection. Data are separated into training, test, and sometimes validation set. The model is trained exclusively on the training data while the model performance and thus the generalization ability of the model is calculated on the test and validation data. Fig. 2.1 illustrates the thus occurring dilemma. Models of high complexity tend to memorize the data which can lead to bad generalization while too low complexity might not be able to reconstruct the true complexity of the problem. The issue of model overfitting is more widely discussed by Hawkins [Haw04].

Model Complexity
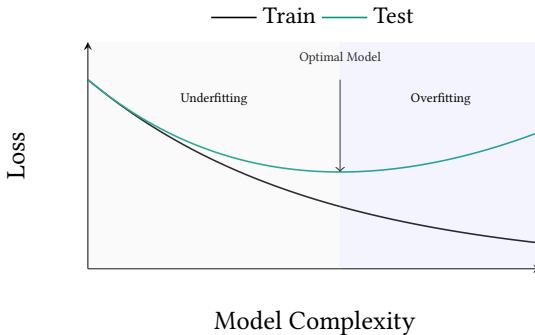
Figure 2.1: Visualization of underfitting versus overfitting dilemma.

Overfitting can occur in all models presented in the following, which is why often ML engineers work together with experts in the relevant subject to quickly build models of representative complexity.

2.4.1 Multiple Linear Regression

The multiple linear regression combines a multitude of features to a prediction via

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x + \epsilon \tag{2.15}$$

where: y Predicted value
x Feature (regressor/predictor variable)
β Regression coefficients
ϵ Residual

The number of features is not limited and can also include combinations of features to represent dependencies. Furthermore, the linear regression only demands linearity in the regression coefficients, not in the regressors. [Mon12]

An estimation of the regression coefficients can be calculated analytically with the ordinary least-squares estimation. This method is based on the minimization of

$$\sum_{i=1}^{n} \epsilon_i^2 = \epsilon^T \epsilon = (y - X\beta)^T (y - X\beta) \quad . \tag{2.16}$$

This leads to the regression parameters of $\beta = (X^T X)^1 X\, y$.

Another approach to estimate the regression coefficients is the Least Absolute Shrinkage and Selection Operation (lasso) [Tib96, Xu08] method. The method is also based on eq. (2.15). However, some regression coefficients are shrunk and some even suppressed to be zero so that the solution is no longer analytically defined but the minimization problem shown in eq. (2.17). This is considered to lead to higher prediction accuracy since it tackles the problems with the original least-square of high variance with low bias. [Tib96]

$$\min_{\beta} = \sum_{i}^{n} \left(y_i - \beta^T x_i\right)^2 + \lambda \, \|\beta\|_1 \tag{2.17}$$

with λ as regularization parameter. Due to its characteristic of building a sparse β, lasso regression can also be thought of as feature selection method [Zho21]. This property also leads to easier model interpretation than the ordinary least-square method [Tib96].

2.4.2 Artificial Neural Networks

Artificial Neural Networks (ANNs) are inspired by the way the human brain functions [Abr05, Kro08]. The first ANN was built by Rosenblatt in 1957 with only one layer of neurons. This is referred to as *perceptron* [Fra57]. An artificial neuron, as shown in fig. 2.2a, receives multiple inputs and combines them linearly with a weight assigned to each of these inputs. The bias term is an additional predefined offset that is added to the neuron independently from the received other input values. Function f_a is referred to as activation function. As indicated by the arrow in fig. 2.2b, a standard feed-forward ANN has

a strict direction for the flow of information. Other network structures might also include information flowing backwards (as e.g., described in [Med01]) . For networks with an arbitrary amount of hidden layers and neurons, it was shown that with the use of unbounded activation functions, ANNs can be used to approximate every possible function [Son17].

Training the network can be done through different approaches. In this dissertation, all training is conducted as supervised learning. This implies that labeled data are applicable and the model receives inputs with correlated labels for the training. For an overview of different learning scenarios refer to [Moh18].

(a) Principle sketch of a single artificial neuron

(b) Principle sketch of a feed-forward ANN
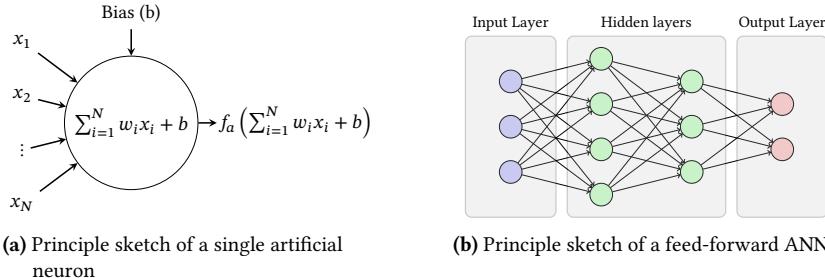
Figure 2.2: Principle sketch of basis element in ANN and standard feed-forward ANN.

Training the model can be done by backpropagation in which the gradient of each weight and bias with respect to the cost function is calculated. These are then updated as

$$W_{lij}^n = W_{lij}^{n-1} + \Delta W_{lij}^{n-1} \qquad b_{lj}^n = b_{lj}^{n-1} + \Delta b_{lj}^{n-1} \tag{2.18}$$

with W_{lij}^n as weight matrix where i and j indicate the number of inputs and the corresponding neuron in the specified layer l. n indicates the number of steps already taken in training. Equivalently b_{lj}^n indicates the bias of neuron j in layer l.

Generally, the necessary ΔW_{lij}^{n-1} to update the weights can be constructed differently and poses an optimization problem in itself. In the most simple form of the gradient descent it is constructed from the partial derivatives of the ANN

$$\Delta W_{lij}^{n-1} = -\eta \frac{\partial C}{\partial W_{lij}^{n-1}} \qquad \Delta b_{lj}^{n-1} = -\eta \frac{\partial C}{\partial b_{lj}^{n-1}} \tag{2.19}$$

where η is the learning rate and C the cost function which indicates the performance of the network. The so far described learning algorithm is referred to as gradient descent. Often, an additional term is added to the weight update

$$\Delta W_{lij}^{n-1} = -\eta \frac{\partial C}{\partial W_{lij}^{n-1}} + \alpha \Delta W_{lij}^{n-2} \qquad \Delta b_{lj}^{n-1} = -\eta \frac{\partial C}{\partial b_{lj}^{n-1}} + \alpha \Delta b_{lj}^{n-2} \tag{2.20}$$

with α as momentum which influences the current weight and bias update in dependence of updates in earlier iterations. This can be especially useful to shorten calculation time and decrease the probability of getting stuck in a local minimum. The backpropagation essentially calculates partial derivatives of the cost function with respect to each weight and bias through the chain rule. Some algorithms include an additional coefficient for the optimization step, the weight decay. This in essence describes a way to avoid large weights by adapting the loss function through e.g. the L2-Norm of the weights [Kro91, Smi18].

The learning rate, the momentum, and the weight decay are so called hyperparameters. These describe parameters that control the learning process and describe the network architecture [Moh18]. As earlier mentioned, the ANN also makes use of activation functions. These are non-linear functions. Otherwise, the ANN would not be able to learn non-linear dependencies since it would simply build a highly complex multiple linear model. In this dissertation, the activation functions displayed in fig. 2.3 are used. However, generally each non-linear function might be applicable [Sha17].
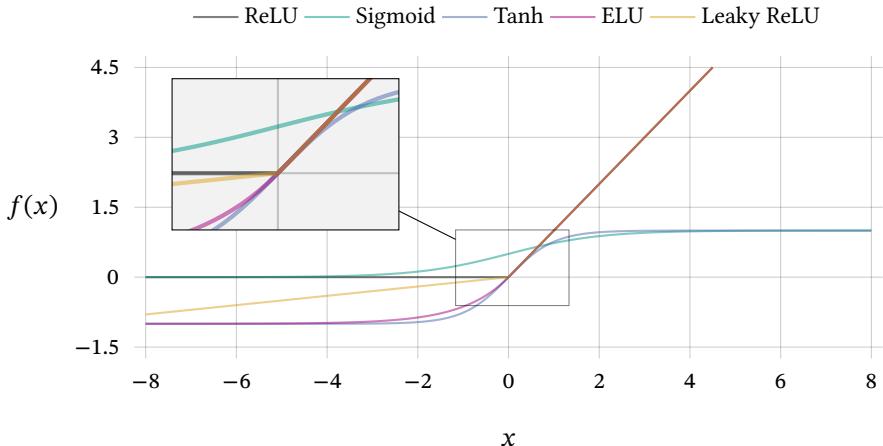
Figure 2.3: Overview of activation functions used in this dissertation.

The functions are mathematically defined as

$$\text{ReLU}: \quad f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.21}$$

$$\text{Sigmoid}: \quad f(x) = \frac{1}{1 + e^{-x}} \tag{2.22}$$

$$\text{Tanh}: \quad f(x) = \tanh(x) \tag{2.23}$$

$$\text{ELU}: \quad f(x) = \begin{cases} x & \text{if } x \geq 0 \\ e^x - 1 & \text{if } x < 0 \end{cases} \tag{2.24}$$

$$\text{Leaky ReLU}: \quad f(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha x & \text{if } x < 0 \end{cases} \tag{2.25}$$

with $\alpha = 0.1$ in fig. 2.3. Training models is typically performed iteratively in epochs. One epoch indicates that the model has seen the whole training set once. Furthermore, training is performed online or in batches. In online learning, after each random sample, an optimization step in weights is taken. For Gradient Descent, this refers to Stochastic Gradient Descent. In batch training, the gradients are calculated for the whole batch of the training data, for Gradient Descent this refers to Batch Gradient Descent. A compromise between

these both approached is reached for mini-batches where the whole data is reorganized in mini-batches of defined size and a step of the optimizer is taken based on each processed mini-batch. In the case of Gradient Descent optimization this refers to Mini-Batch Gradient Descent (MBGD). As necessary gradients and loss values, the average of the batch is used. [Abr05, Ian16]

The terms online and stochastic are sometimes used interchangeably, however online is more commonly used if the data is continually created during the training process. It can be said that larger mini-batches lead to more accurate estimations of the gradients during training, however using the whole data as batch also demands much more hardware resources. [Ian16] Other literature argues, that online training leads to faster convergence due to multiple steps being taken in each epoch and the advantage of this since it allows for curved gradients in the data set compared to straight gradients in batch learning. [Wil03]

As cost function, which is necessary for training, arbitrary functions can be defined. A common function for regression tasks is the Mean Squared Error (MSE) or the Mean Absolute Error. Through the squared error, MSE punishes outliers more strongly than the Mean Absolute Error.

Even though ANNs can approximate any function, they are not suitable for all problems. This applies in particular to image processing and pattern recognition. Standard ANNs are fully connected. This combined with the thought of processing every single pixel of an image as input node, leads to computationally inefficient networks. [OSh15] This issue is overcome with CNNs presented in the following section.

2.4.3 Convolutional Neural Networks

The general structure of CNNs is rather similar to that of ANNs. However, not all layers are fully connected. [OSh15] In contrast to standard ANNs no feature extraction needs to be performed. CNNs can find them themselves. It is thereby based on human vision [Hub62].

Standard CNNs consist out of three types of layers:

- Convolutional layers

- Pooling layer

- Fully connected layers

The convolution layer receives a matrix and applies a mathematical convolution on the matrix. This is represented by sliding a so called kernel over the image. Fig. 2.4 demonstrates the process with corresponding output. The kernel is a learnable entity so that the model learns itself how to convolve the input for the best data representation. The shape of the kernel is generally predefined as model hyperparameter.
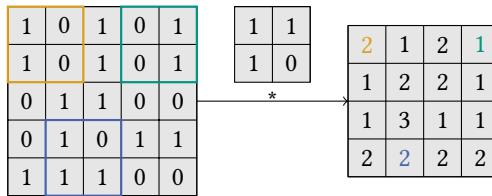
Figure 2.4: Principle sketch of a 2D convolution with 2x2 kernel on 5x5 input data. Different colored areas in input data display the interaction of the kernel, input data, and the identically colored output.

CNNs can be used for any data type but they are especially suitable for pattern recognition. In this dissertation, they are used for image processing, this is why following explanations also refer to an image input. Like the convolutional layer, the pooling layer is based on a kernel that slides over the input data. The essence of pooling layers is to reduce the dimension of the input by combining multiple pixels into one, as indicated in fig. 2.5 for average pooling or max pooling. Both, the convolution as well as the pooling layer, demand additional parameters: stride and padding. Stride defines how many input data points the kernel slides over for each step. Padding is necessary to specify how to process the edges of the data. As illustrated in fig. 2.4, the 5x5 input dimension processed with a 2x2 kernel and no padding leads to a reduction of the output dimension to 4x4. This is a result of the necessity for the

kernel to be fully within the range of the input data dimension. To also convolve the border areas of the input, padding can be applied, which essentially enlarges the input dimension. For example, the common zero-padding of one enlarges the 5x5 data to 7x7 with the additional values all being zero. It should be mentioned that the kernels do not have to process directly connected pixels. Through the concept of dilation, kernels can cover larger spatial areas than their own size and for example only process every second pixel. Via deformable convolution, it is furthermore possible to enable the network to train its own sampling location in kernels [Jif17].
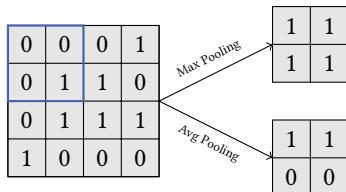
Figure 2.5: Principle sketch of average (avg) and maximum (max) pooling with 2x2 kernel, indicated in blue, on 4x4 input data without padding.

The fully connected layers have the same tasks as in ANNs. They map the flattened activation maps to actual labels.

All the above-mentioned layers can be applied on 1D, 2D, and 3D data. The input of a CNN is defined through height, width, and depth. The depth can be thought of as channels in images. [OSh15]

The whole CNN is constructed through multiple layers in sequence. An example of how the spatial dimensions for an image input could look like is shown in fig. 2.6. The figure especially illustrates the common approach of increasing the number of channels in each convolution. This is seen in the increasing width of the spatial data. In processing terms this results from multiple kernels that process the same spatial area. A stack of kernels is referred to as filter. Activation functions as presented in section 2.4.2 are also included in CNNs. The representation achieved after each layer is referred to as activation or feature maps. In CNNs not each neuron is connected to all neurons in

the preceding layers. They rather correspond to a fixed area which is called the receptive field. Also, kernels are not calculated for each receptive field individually, they are shared along a layer, which means the weights are also shared for many connections. Those principles greatly reduce the necessary parameters and therefore help in faster convergence. Lastly, pooling layers also reduce spatial dimension and with it the number of trainable parameters. [OSh15, Ket17, Li21]
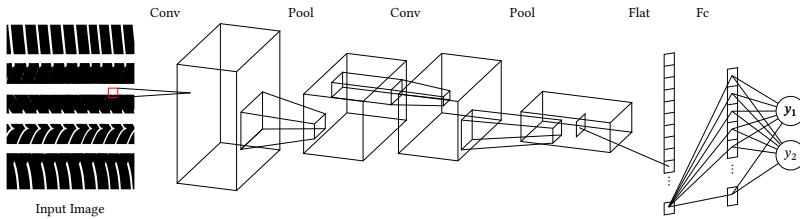
Figure 2.6: Schematic model representation of CNN structure.

For a further overview of models in deep learning, the interested reader is referred to [Sch15, Ian16, Alo19, Pou19, Don21].

2.4.4 Bayesian Hierarchical Modeling

In contrast to the Frequentist approach, which defines probability as relative frequency of a reoccurring event, the Bayesian approach interprets probabilities as measure of belief. The Bayesian approach does not try to predict single regression coefficients but rather a distribution of these coefficients. This implicitly allows the model to state uncertainty in the prediction and with it illustrates a convenient way to estimate model accuracy.

Bayesian linear regression models can be described through eq. (2.15), only that the coefficients β_i are probability distributions and not fixed values. For this, Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.26}$$

is used, which allows for a calculation of the conditional probability of $P(A|B)$ if the conditional probability $P(B|A)$, and the prior probabilities $P(A)$ and $P(B)$ are known. The prior probability $P(A)$ represent the probability of event A occurring without any constraints. The conditional probability $P(A|B)$ represents the probability of observing A under the premise that B has also occurred. In the context of machine learning models this can be rewritten as

$$\overbrace{P(\theta|D, M_j)}^{\text{Posterior Distribution}} = \frac{\overbrace{P(D|\theta, M_j)}^{\text{Likelihood}} \overbrace{P(\theta|M_j)}^{\text{Prior}}}{\underbrace{P(D|M_j)}_{\text{Evidence}}} \tag{2.27}$$

where: θ Parameters of the model
M_j Specific model j
D Observed data

[Bar12, Gha15, van21]. The idea in Bayesian models is to approximate the posterior distribution through changes in the parameterization of the likelihood to create a posterior distribution that accurately represents the observed data. The prior defines the a-priori believe about the model parameters, this describes the assumed range and distributions of the parameters before seeing the data. The evidence can be thought of as normalization factor and describes a marginal likelihood and can be calculated through integrating over all possible values of the parameter range for a specific model

$$P(D|M_j) = \int_{\theta \in \Theta} P(D|\theta, M_j) P(\theta|M_j) \, d\theta \quad . \tag{2.28}$$

Bayesian statistics are constructed in three steps: [Was04]

1 Defining prior distributions $\pi(\theta)$ for each parameter θ. These represent the belief or knowledge about parameter θ before analyzing data.

2 Defining the likelihood $P(D|\theta, M_j)$ which represents how likely it is to receive the data with the condition of parameter θ for model M_j. This

represents a statistical model which allows for data generation with the assumption of parameters θ and model M_j.

3 Use available data D to calculate the posterior distribution and update prior beliefs.

Inference can be achieved in different ways. Two very common ones are the variational inference and the Marcov Chain Monte Carlo (MCMC). In the variational inference the posterior distribution is approximated through a combination of simpler distributions. For these, the inference algorithm optimizes the parameters to minimize a loss function. The MCMC embodies an approach to estimate the true posterior of any distribution by combining Monte Carlo sampling and empirical probability estimation through the Markov Chain. The variational inference demands lower computational effort but also results in a generally less accurate posterior than the MCMC inference. [van21] A Markov Chain is a sequence of elements in which the conditional distribution of a consecutive element is dependent only on the immediately preceding element. [Bro11] The higher accuracy of MCMC comes at high computational cost, which is why in cases where computational cost is more important than accuracy, variational inference might be preferred over MCMC [Sal15].

The priors are generally divided into weakly informative, informative, and diffuse priors. [van21] The latter are also referred to as non informative, vague, or flat priors [Gel13]. These basically categorize how much information the model received through the priors. An informative prior is therefore e.g. a Gaussian distribution with a mean close to the expected value and a small variance. Priors like these allow for information that might not be included in the data but has been achieved by previous studies. [van21]

Bayesian models are also referred to as hierarchical models since Bayes' theorem induces a hierarchical structure through the multiplication of a conditional distribution with the distribution of the parameter on which the preceding is conditioned [Lyn07]. The hierarchical structure also allows for adding information on different levels to the model. Data can be grouped to build a compromise of pooled and unpooled models. [Vee23] Unpooled models separate the data into different fragments and train a model for each of these data

fragments independently. In case of pooled models one model is trained for the entirety of the data. [Gel13] Fig. 2.7 illustrates the concept of a hierarchical Bayesian model, in which data are separated in i fragments, yet the model parameters θ_i are connected through a prior distribution. The prior π can also be specified through overlying hyperpriors, these would then describe μ_π and σ_π in fig. 2.7 through a distribution. [Gel13] The whole process of model design and evaluation is also described in [Gel20].

$$\pi(\mu_\pi, \sigma_\pi)$$

$$\theta_1 \quad \theta_2 \quad \theta_3 \quad \cdots \quad \theta_i$$

$$Y_{j1} \quad Y_{j2} \quad Y_{j3} \quad \cdots \quad Y_{ji}$$
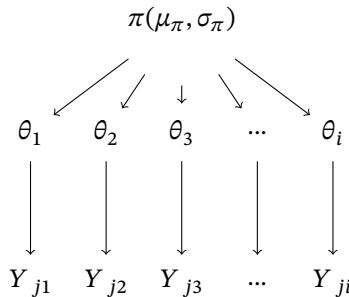
Figure 2.7: Principal sketch of a hierarchical Bayesian model. Adapted from [Gel13].

2.4.5 Cross-Validation

As described above in this section, model selection is an important topic in ML. This occurs out of dilemma between over- and underfitting. The model performance is estimated with an arbitrary metric on at least two types of data: The training and the test data. Training data implies, that the included data points are used in the training process. The correlating training loss indicates how well seen data are represented by the model. The test data are explicitly left out of the training process, therefore the test loss estimates the true prediction error. The split in the data can have a major impact on the recorded losses, which is why methods to divide the data are an important research topic. A common approach is the cross-validation in which the model

is trained multiple times for different data splits and the overall model performance is recorded as average performance of all trained models. In the k-fold cross-validation, the whole data is divided into k approximately equally large subsets. These are built through random sampling from the entirety of the data without replacement so that no data point exists in more than one subset. The training data are assembled through combining $k - 1$ data packages, the omitted data are used as test data. Fig. 2.8 illustrates the iterative process. The special variant of the method is reached for $k = n$ with n as number of data points. In this case only one data point is left out, which is why this method is referred to as Leave-One-Out Cross-Validation (LOOCV). For larger data sets this is however rather computationally costly. In some cases, the division of the data into subsets is not performed randomly but rather in a way to ensure equivalent populations of each subset.
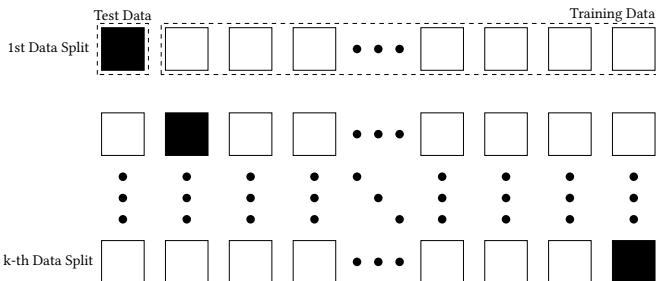
Figure 2.8: K-fold cross-validation principal sketch based on [Ber19]. White square indicate batch used in training data and black square indicate batch used as test data.

2.4.6 Hyperparameter Optimization

Parameters that are not determined through the training process are referred to as *hyperparameters*. These parameters are used as input for the learning algorithm to define the architecture of the model as well as the learning process itself. [Moh18] Included parameters for ANNs and CNNs are, but not limited to, number of hidden layers, number of neurons in each layer, number and size of kernels in the convolutional layers, pooling layer variation and kernel size,

learning rate, optimization algorithm, activation function, batch size, dropout rate, weight decay, loss function, and number of epochs. Overall, these parameters define the model architecture, regularization, and the optimization process [Feu19]. These parameters can substantially impact the model performance but yet bear the challenge of being chosen optimally. Especially in Deep Learning (DL) it is difficult to estimate these parameters since it is also not straightforward to estimate the necessary model complexity.

A standard approach for hyperparameter optimization is to split the data into three parts: Train, validation, and test set. The validation set is solely used to optimize the hyperparameters, while the test set is used to estimate the overall performance of the optimized structure and learning process. [Moh18]

Due to the high variance in parameters, recent research investigates different optimization algorithms to let the system define these themselves. Some of these approaches are explained in [Feu19].

2.5 Explainable Artificial Intelligence (XAI)

One problem of ANNs and CNNs, especially if they are designed as DL models, is that they are black box models [Gui19]. This basically means, other than in multiple linear regression models (see section 2.4.1), the importance of features, also known as feature attribution, is not directly available. This is unpleasant since it prohibits users to gain further understanding of the models predictions and also decreases the amount users trust in them. This also makes it more complex to estimate if actual correlations of the problem are learned or if rather scenarios like shortcut learning [Gei20] occurred. In shortcut learning, the model acquires decision rules from the training set that are present in the data but may not necessarily correlate with the actual problem. As a consequence, these rules lead to inaccurate predictions when applied to other data sets. [Gei20]

These aspects lead to a recent development of many algorithms in Explainable Artificial Intelligence (XAI). The strong drive in demand for explainability can be divided into four major motivations: [Ada18]

1. Justification: implies that an erroneous prediction or decision can be ruled out

2. Control: gives deeper insight in the models decision process and leads to easier implementation of changes

3. Improvement: deeper insight allows faster implementation of new knowledge

4. Discover: user can gain additional knowledge on the matter through the model

Kokhlikyan et al. classify the algorithms in XAI by two metrics, first being the depth of explanation and the second as the underlying algorithm. The depth is divided into primary-, neuron-, and layer attribution, from which primary is commonly described as feature attribution. The classification by algorithm is divided in gradient, perturbation, and others. [Kok20] A similar classification via core algorithmic into gradient and perturbation based is done by Das et al. [Das20]. While the first one makes use of the gradients calculated with backpropagation to explain the model, the latter induces small changes in the input data to investigate the correlated model response [Das20]. In this dissertation, one perturbation based and one gradient based method are applied. These are described in the following sections. For a general overview of XAI, the reader is referred to [Ada18, Gui19, Iva21].

2.5.1 Integrated Gradients

The method Integrated Gradients (IG) was first introduced in 2017 as a gradient based method for XAI. The idea behind IG is to integrate the gradients of the model along a straight line that connects two arbitrary input vectors.

The IG can be described through

$$
\mathrm{IG}_i^\xi(x) = \int_{\zeta=0}^1 \frac{\partial f(\xi(\zeta))}{\partial \xi_i(\zeta)} \frac{\partial \xi_i(\zeta)}{\partial \zeta} \qquad .
\tag{2.29}
$$

where: f Function of model
        i Dimension i of vector
        ζ Interpolation integrator
        $\xi(\zeta)$ Linear path between two points

Substituting the proposed linear path between two point defined by $\xi(\zeta) = x' + \zeta(x - x')$ and approximating the integral trough the Riemann approximation leads to

$$\text{IG}_i^{\text{approx}}(x) = (x_i - x_i') \sum_{k=1}^{m} \frac{\partial f\left(x' + \frac{k}{m}\left(x - x'\right)\right)}{\partial x_i} \frac{1}{m} \qquad . \qquad (2.30)$$

where: x Input sample
        x' Baseline
        k Interpolation step
        m Number of steps in Riemann Integral

[Sun17] proof axiomatic that the attribution method satisfies the following principles:

- Completeness: equality between the sum of attributions and the actual change in output value from the baseline to investigated input

- Sensitivity: a change in a feature that results in changes of the output, for all other features kept constant, results in an attribution unequal to zero. Similarly if this change in feature value does not lead to changes in the output, its attribution is zero

- Implementation invariance: two individually designed models which result in equivalent output for equivalent input assign identical attributions to their features

- linearity: combining models linearly with weighting factors leads to the weighted sum of the attribution of the singular models as combined model attribution

- Symmetry-preserving: two variables that are interchangeable for each of their values and thus do not cause any changes in the output are

> assigned the same attributions if their input and baseline values are
> identical

An advantage of IG is the application which does not need any changes in the ML model since it is performed after model training. As baseline the authors suggest using a value with prediction close to zero, which they consider to be a black image for image processing. [Sun17] However, it was pointed out in a later comment that the method is basically blind to the defined baseline. A black image as the baseline therefore means that the algorithm is not influenced by black pixels when calculating the IG. [Sun18]

This led to a more profound investigation of different baselines by Sturmfels et al. For a specific image, four approaches are investigated: The maximum baseline which describes an image that is the furthest away from the current image in each pixel value; the blurred baseline which blurs the current image; the uniform baseline which represents an image drawn from uniform distribution of the whole pixel value range; the Gaussian baseline, which adds noise to every pixel of the current image drawn from a Gaussian distribution. Especially for the latter two approaches, multiple draws from the distributions and averaging over these is also considered by the authors. Even though, the baseline leads to different attributions on different data, the authors conclude that there is no overall best performing baseline since it always makes assumptions about the distribution in the data and the explicit understanding of missingness of the specific problem. [Stu20] Even though, the definition of the baseline can lead to difficulties in interpreting the attributions, IGs have advantages over methods that consider gradients of single inputs without interpolation between baseline and input. This is a result of saturation effects in DL which represents a sort of threshold above which no changes in the predictions occur for additional changes in the input value. [Stu20] Still, the saturation effect does impact the attributions computed through IG. This is shown by the alternative approach of splitting the regions of IG into two parts, one above and one below a predefined value. This value is defined through a threshold in absolute maximum value of the predicted output, so that the overall interpolation space is separated in saturated and not saturated areas. The authors conclude on basis of their case study that calculated attributions

with IG might be dominated by saturated areas. They therefore recommend an adapted, two sided version of the method. [Mig20] It should be noted that this might also be a result of the chosen baseline, reported as gray, which leads to bright colorful pixels having very different values which might lead the $x - x'$ multiplier in eq. (2.29) to overestimate the importance of the saturated area.

Another adaption of the method is achieved through adaptively changing the path along which the gradients are integrated. In the concept of Guided Integrated Gradients only some features are changed each step. These are chosen as the features with the lowest absolute partial derivatives. This avoids points with anomalies to induce noise to the attributions. [Kap21]

2.5.2 Shapley Additive Explanations

Shapley values were first introduced by Lloyd S. Shapley in 1953 [Sha53] as co-operative game theory. They try to rank each player by their impact on the output of the game. This is achieved by averaging the marginal contribution of a player on each coalition over the total number of coalitions. A transfer from game theory to ML is achieved through considering each input feature as players.

They can be calculated via

$$\phi_j = \sum_{Z \subseteq \{1,\dots,n_F\} \setminus \{j\}} w_Z(Z) \Big[\overbrace{\nu(Z \cup \{j\}) - \nu(Z)}^{\text{marginal contribution of j to Z}} \Big] \qquad (2.31)$$

$$w_Z(Z) = \frac{|Z|!(n_F - |Z| - 1)!}{n_F!} \qquad (2.32)$$

where: ϕ_j Shapley value of feature j
           Z Coalition without j
           $Z \cup \{j\}$ Coalition with j
           n_F Overall number of features
           ν Function measuring the usefulness of the coalition (also called characteristic function)
           w_Z Weight function for probability of entering in coalition Z

[Sha53, Lip01, Mol22] This also includes the concept of *usefulness* introduced by Darlington, who defines it as the decline in R^2 after excluding the specific variable in a multiple linear regression. [Dar68]

Eq. (2.31) is constructed for linear models including multicollinearity and demands a multitude of models to be trained to gain insight into the model performance for only subsets of features. By approximating the equation by sampling this can be achieved much less computationally heavy. Lundberg et al. introduce multiple new approximation methods which are more efficient or model specific. [Lun17]

They are both derived from the sampling method presented in 2014 by Štrumbelj et al. to approximate Shapley values which is explained below. A huge advantage of that method is the applicability on any model since no changes are necessary due to its sole investigation of the relation between input and output. In case of the sampling, it is important to ensure that the distribution of each feature is independent of another. The process is abbreviated from the Monte Carlo integration. [Štr14]

To calculate the contribution of a feature j for a data point x_1, the corresponding Shapley values are calculated a user defined n times and averaged with respect to n

$$\phi_j = \frac{1}{n} \sum_{i=0}^{n} \phi_{j,i} \quad . \tag{2.33}$$

take corresponding values from x_1	←	features preceding j in $\mathcal{O}$	→	take corresponding values from x_1
take corresponding value from x_2	←	feature j	→	take corresponding value from x_1
take corresponding values from x_2	←	features succeeding j in $\mathcal{O}$	→	take corresponding values from x_2

d_2 d_1
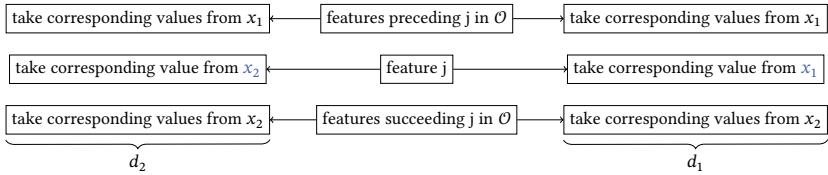
Figure 2.9: Construction rule of new data points to approximation of the Shapley values. (Adapted from [Štr14])

The Shapley value of each iteration is calculated as shown in eq. (2.34) as the difference in the models prediction for two newly constructed data points d_1 and d_2.

$$\phi_{j,i} = f(d_{1,i}) - f(d_{2,i}) \tag{2.34}$$

These new data points are constructed for each iteration as combination of the data point under investigation x_1 and an independently sampled data point x_2 from all possible data points. The construction is done with a randomly drawn perturbation $\mathcal{O}$ of all features. The perturbation defines how the new data points are constructed during each iteration, the associated rule is displayed in fig. 2.9. [Štr14] The exact Shapley values as defined by Shapley also fulfill some desirable properties: [Sha53, Roz22, Mol22]

- Efficiency, which states that the sum of all Shapley values adds up to the overall value of the coalition (similar to completeness in section 2.5.1)

- Symmetry, if two features contribute identically to all coalitions, they get assigned the same Shapley value

- Dummy, a feature that does not impact any coalition, gets assigned a Shapley value of zero

- Linearity, this implies that scaling the output values also scales the Shapley values with the same factor

3 Data Acquisition

This chapter describes the processes used to acquire data necessary for the TRN modeling, gives an overview of the performed experiments, and presents the tire tread pattern image processing steps.

3.1 Acoustical Methodology and Measurement Process

All measurements included in this research are executed at the testing facilities of Mercedes-Benz (MB) situated in Immendingen, Germany. The measurement track satisfies the specifications described in ISO 10844 [Deu23]. Fig. 3.1 visualizes an according test track with its specific dimensions and point labels. Measurements are performed with the PAK Pass-By Software versions 3.2 and 3.3 from Müller-BBM VibroAkustik Systeme GmbH.

During each run, the SPLs on each side of the vehicle are recorded, as well as the vehicle speed, air temperature, road temperature, wind speed, wind direction, and humidity. The vehicle speed in combination with light barriers, which start the measurement when triggered by the front of the vehicle, is also used to calculate the vehicle position on the test track. All environmental values are always recorded at approximately the same position, which is located diagonally the furthest away from the center point of the test track. At this point, also the road temperature is measured through the weather station. The microphones are placed at a height of 1.2 m, at a distance of 7.5 m to the center line at the 0 m location of the track according to ISO 10844.
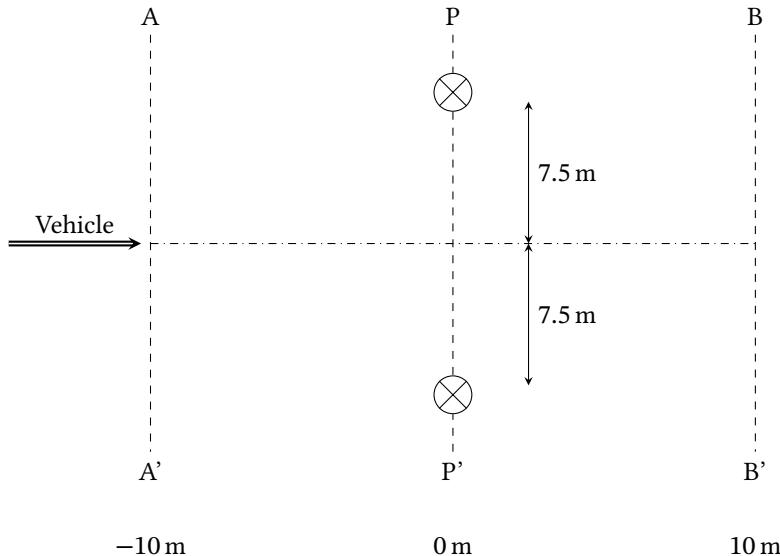
Figure 3.1: Principal sketch of the measurement track. [Leu22]

To combine both UNECE regulations (no. 51 and no. 117), speeds from $40\,\mathrm{km\,h^{-1}}$ to $90\,\mathrm{km\,h^{-1}}$ are included in this study. Each speed is recorded twice, which is theoretically once from AA' to BB' and once from BB' to AA'. However, if a run of these two measurements failed one of the other criteria, multiple runs from the same direction are accepted as well. The speed is increased in increments of $5\,\mathrm{km\,h^{-1}}$ leading to at least 22 measurement per tire. All runs are conducted according to CB method in idle mode. Tires are always mounted in sets of four, no mixed dimensions are used to ensure that a whole measurement series of one tire can be directly connected to one specific tire size, brand, and tread pattern. The inflation pressure is adapted to 2.5 bar for each measurement series after a warm up period of 15 min. This ensures that the pressure is constant during the whole measurement series, making further analysis easier. The measurement speeds are chosen randomly to avoid conditioning the tire through constantly increasing or decreasing the vehicle speed. All tires are driven for approximately 200 km for preconditioning before the measurements.

To ensure the dominant noise source is the tire and not exhaust, engine or background noises, measurements are only conducted with electric vehicles, an acoustically encapsulated vehicle, or in the case of one tire a vehicle with very quiet combustion engine which is not audible while idling. The different tire load resulting through different measurement vehicles is not compensated. The background noise is always at least 10 dB(A) below the maximum recorded SPL of each run. Measurements with higher wind speeds than $5 \, \mathrm{m \, s^{-1}}$ are not recorded. The air temperature is also in the range of $5 \, ^\circ\mathrm{C}$ to $40 \, ^\circ\mathrm{C}$ and the road temperature in the range of $5 \, ^\circ\mathrm{C}$ to $50 \, ^\circ\mathrm{C}$, as demanded by UN regulation no. 117.

Additional to acoustical measurement, the rubber hardness as Shore A as well as the tire profile depth is measured. For the rubber hardness, the HPE III device from Bareiss Prüfgerätebau GmbH is used. This allows measurements of the surface temperature as well as the hardness. The shore hardness is measured once in the shoulder region and once in the center region for each tire. For all four identical tires the average of these measurements is used. The measurements are conducted within three months of the acoustical measurements to ensure that the surface temperature is around $20 \, ^\circ\mathrm{C}$ and the tire did not deteriorate. Profile depth is measured with a caliper in a circumferential groove for each tires. The average of four tires of these measurements is used as model input.

3.2 Tires and Vehicles in Experiments

In total this study includes measurements of 42 summer tires, from 14 different tire manufacturers. These include sizes ranging from tire nominal width of 205 mm to 285 mm, aspect ratios of 30 to 65, and rim diameters of 16 in to 22 in. Table 3.1 presents a full design of experiments as well as the ratio between the load applied by the vehicle and the load index. The tires are herein sorted according to their usage in later presented analysis. This means row one corresponds to data set 0 or tire 0 and row 42 corresponds to data set 41 or tire 41. The vehicles 243 (EQA) and 293 (EQC) are both electric SUVs, the 212 is an older E-Class that was specially modified for tire measurements so

that its engine and exhaust are not audible. The 167 (GLS) is a vehicle with combustion engine. However, the used version is very silent so that the exhaust and engine sound are not audible while idling and the tire noise can be considered as the dominant noise source.

Table 3.1: List of all tires included in measurements with associated measurement vehicle. The load ratio column puts the load of the measurement vehicle and the load ratio in relation.

Vehicle	Manufacturer	Profile	Dimensions	Load ratio
212	Dunlop	Sport BlueResponse	225/55R19 95V	0.76
212	Goodyear	Wrangler HP	215/60R16 95H	0.76
212	Michelin	Energy Saver MO	205/65R16 95V	0.76
243	Bridgestone	Alenza 001 MO	235/45R20 96W	0.72
243	Continental	Premium Contact 5 MO	225/55R17 97Y	0.70
243	Continental	Premium Contact 6 MO	225/45R18 95Y	0.74
243	Continental	Ultra Contact	215/60R17 96H	0.72
243	Continental	4x4 Contact	235/50R18 101H	0.62
243	Fortuna	Ecoplus UHP	235/45R18 98W	0.68
243	Goodyear	Eagle F1 Asymmetric5 MO	225/40R19 93Y	0.78
243	Goodyear	Efficient Grip	255/40R18 95Y	0.74
243	Goodyear	Excellence	245/45R19 98Y	0.68
243	Michelin	Pilot Sport4 SUV MO	255/45R20 105W	0.55
243	Michelin	Primacy4 MO	225/45R18 95Y	0.74
243	Nexen	N Fera Sport SU2 MO	225/55R18 102Y	0.60
243	Pirelli	P Zero Nero GT	235/35R19 91Y	0.83
243	Pirelli	Cinturato P7 MO	235/55R18 104T	0.57
243	Pirelli	Scorpion Verde MO	235/55R19 101V	0.62
243	Viking	ProTech NewGen	225/50R17 98Y	0.68
243	Vredestein	Ultrac MO	225/55R18 102Y	0.60
243	Hankook	Ventus S1 Evo 3 MO	225/55R18 102Y	0.60
293	Dunlop	Dueler H/P Sport MO	235/55R19 101V	0.76
293	Bridgestone	Potenza S001 MO	245/45R19 102Y	0.74

Vehicle	Manufacturer	Profile	Dimensions	Load ratio
243	Bridgestone	Turanza T005 MO-S	255/40R20 101Y	0.76
293	Continental	ContiSportContact 5 MO1	255/50R19 103Y	0.71
293	Continental	EcoContact 6Q MO	235/55R19 105W	0.68
293	Cooper	Zeon Crossrange MO	255/45R20 105H	0.68
293	Dunlop	SP Sport Maxx GT	275/30R20 97Y	0.86
293	Goodyear	Eagle F1 SuperSport	265/35R20 99Y	0.81
293	Kumho	Ecsta PS91 MO	245/45R18 100Y	0.78
293	Michelin	ePrimacy MO	245/45R19 102Y	0.74
293	Michelin	Latitude Sport 3 MO	235/60R18 103V	0.71
293	Michelin	Pilot Sport 4 SUV MO	235/45R21 101Y	0.76
293	Michelin	Pilot Sport 5 MO	245/40R19 98Y	0.83
293	Michelin	Pilot Sport Cup 2 MO1	265/35R19 98Y	0.83
293	Michelin	Pilot Sport 4 SUV MO	255/40R21 102Y	0.74
293	Pirelli	P Zero (PZ4) MO	235/45R20 100T	0.78
293	Pirelli	Powergy MO	255/45R19 104Y	0.69
293	Pirelli	P Zero Corsa Asimmetrico2	265/30R19 93Y	0.96
293	Dunlop	SP Sport Maxx GT MOE	245/35R20 95Y	0.91
293	Yokohama	Advan Sport (V105) MO	235/60 R18 103V	0.71
167	Continental	Premium Contact 6 MO-S	285/45R22 114Y	0.55

3.3   Data Preprocessing

While some ML models can deal with alphanumeric values, in standard ANNs, data must be converted into numerical values. This is due to the working principle of ANNs which relies on applying mathematical operations on the data. Other algorithms, as e.g., Decision Trees [Rok05, Kin08], can also receive alphabetical input and progress these through their structure.

One method to transform non numeric values into linear numeric space is the one-hot encoding. Thereby the values are transformed into a matrix of shape

NxN for a variable with N different values and each value N_i is represented by a single 1 in row i and column i, while all other rows in column i are filled with zeros. This is illustrated in eq. (3.1) for four values (Dummies) of a specific feature.

$$
\begin{array}{cccc}
\text{Dummy1} & \text{Dummy2} & \text{Dummy3} & \text{Dummy4}
\end{array}
\\
\left(
\begin{array}{cccc}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}
\right)
\tag{3.1}
$$

Other encoding methods could include hashing functions as presented in [Wei09]. Some more encoding schemes are listed in [Pot17].

An additional step in data preprocessing includes standardizing or normalizing. This is a preprocessing step to avoid overweighting features with larger value ranges [Han11]. Al Shalabi et al. provide an overview on different normalization techniques [Al 06]. One of these is the z-score normalization

$$
x_i^{\text{scaled}} = \frac{(x_i - \overline{x})}{\sigma_x} \quad .
\tag{3.2}
$$

The z-score normalization refers to the *StandardScaler* in Scikit-learn [Ped11]. The normalization is performed on each feature individually. The z-score normalization in essence removes the mean of the data and scales the value range to unit variance. Even though data normalization is very common, research investigating the impact on the prediction accuracy for a specific use case with different ML algorithms found no normalization method that leads to the highest accuracy in all algorithms [Ahs21].

3.4 Digitalization Method

In this research two different methods to extract tire images are used. These can be used as input for a CNN and to extract relevant tire tread pattern features. These two methods are briefly explained in the following sections.

3.4.1 Photogrammetry - Structured Light

Digital representations can be accomplished through photogrammetric methods, which include all methods that enable extracting measurement data as length, width, and height from pictures [Mik01, Lin06].

More advanced processes are not limited to extracting some dimensions but allow the reconstruction of a full 3D tessellation model. For this purpose, some commercial as well as open source pipelines (as put together in [Rad22]) are already available. In a previous research, the usability of two such systems is investigated and they are compared on the basis of some acoustical feature extraction. This investigation showed, that the extracted features through the open source pipeline and a commercial tool do not vary gravely. However, the commercial tool is faster in computational time and in the setup, including taking the necessary pictures. [Leu23].

In this research the commercial system ATOS III Triple Scan from Carl Zeiss IQS Deutschland GmbH (former GOM GmbH) is used. The system uses two cameras and one projector. Due to the different orientation of the cameras, some points of the image are recorded multiple times, therefore making the 3D transformation more stable and allowing for deeper pouches to be visualized. [Car23] The photogrammetric process resulting out of the combination of at least one camera and a projector is referred to as structured light approach [Bel99]. The projector stabilizes the process even more, since it allows for an easier matching of different images through the projected light pattern. This results in the possibility to connect every pixel in the images with a unique codeword, making it easier to find corresponding points while overlapping multiple images in the reconstruction process. This is referred to

as encoding. [Bel99] For the reconstruction, the relative orientation of camera and projector must be known [Heh20]. The ATOS III Triple Scan projects multiple different encodings with phase difference of blue light on the objects surface. Due to the lower wavelength of the blue light bad lighting and shiny objects can be better compensated. [Car23] The system outputs tire models as presented in fig. 3.2.
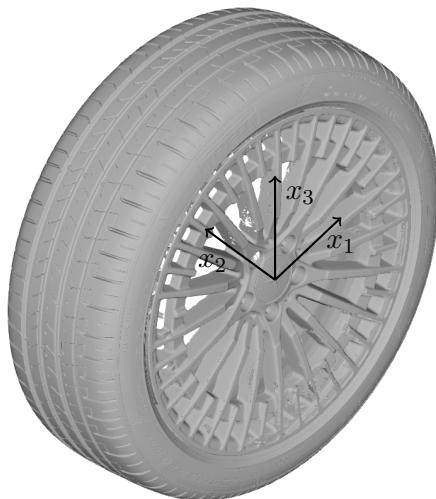
Figure 3.2: Extracted tire model with structured light approach.

For later usage as input for the ML models the tread pattern is transformed to 2D along the circumferential direction. This leads to an image as presented in fig. 3.3 with only the rubber in contact with the road included. In the following these images are referred to as tire circumference images.
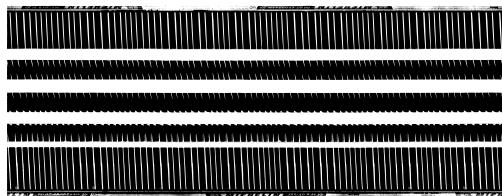
Figure 3.3: Tire circumference image extracted from digital tire model. Image shows whole circumference of a tire. Scales in circumferential and transverse direction are not identical.

The process of extracting these images is explained in depth in [Leu23].

3.4.2 Contact Patch Extraction

The contact patches of tires can be easily extracted with the FUJIFILM Corporation Prescale pressure measurement film [Fuj11]. These are available in different pressure ranges and therefore come as two-sheet and mono-sheet variation. Since in this investigation the purpose is to capture the whole contact area and not the specific pressure distribution, high pressure sensitivities are more useful. The herein specified pressure range, is from 0.05 MPa to 0.2 MPa.

Fig. 3.4 illustrates the composition of such a two-sheet low pressure foil. One sheet includes microcapsules that burst under the predefined pressure and release a liquid dye. The second sheet has developer material as coating, which reacts with the dye, coloring the area under pressure red. By using a series of microcapsules of different sizes that burst under different pressures, a color gradient in correlation with the existing pressure gradient is achieved.
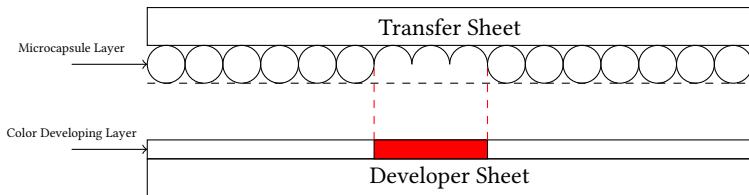
Figure 3.4: Principle of FUJIFILM Corporation Prescale two-sheet type adapted from [Fuj11, FUJ23].

Fig. 3.5 shows the setup to capture the tire tread patterns. The used measurement device consists of three air bellows which allow for a precise adjustment of the tire load. The chosen tire load refer to the actual load during the acoustical measurements. These are 510 kg, 525 kg, 625 kg, and 650 kg for the measurement vehicles 243, 212, 293, and 167. All tires except those with 16 in rim diameter are mounted on the measurement device for the use of the pressure foils. The 16 in rim tires are mounted on the actual measurement vehicle and load is applied by releasing the vehicle from an auto hoist. Between the thick metal sheet which is placed on the ground and the aluminum foil, which is placed on said metal sheet, some mechanic grease is applied to allow the aluminum to move while pressure is being applied. The pressure foil is fixed on the aluminum sheet so that no relative movement between these occurs.

Figure 3.5: Process of creating tire tread patterns with FUJIFILM Corporation Prescale pressure measurement film.

The colored foils are afterwards scanned with 600 dpi to receive digital representations for further processing. Fig. 3.6 shows exemplary images obtained by this process.
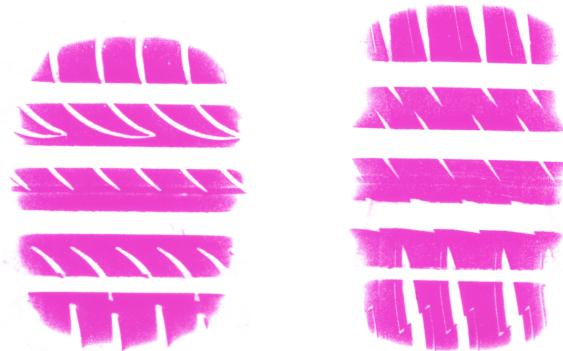
Figure 3.6: Exemplary scans of tire tread visualised with FUJIFILM Corporation Prescale pressure measurement film.

3.5 Image Processing

The tire circumference image, described in section 3.4.1, should have the same pixel size to further process them in a CNN. This however leads to a loss of information if not preprocessed before usage in the CNN. Some of the tires have different dimensions but the same image sizes. This results in a tire with e.g., 205 mm width being scaled differently than a 285 mm wide one. Structure-wise, this is not supposed to change the orientation of grooves, but it makes it impossible for the CNN to extract features as e.g., groove widths correctly.

The easiest approach is to not change the images at all, and assume that more complex models which combine CNN and ANN will learn the correlation between the tire nominal width as input and the different image representations by themselves. Nevertheless, this is not guaranteed at all and therefore not assumed to be the best approach and also is not possible in the investigation

of a single CNN. An alternative approach correlates the image height to a de-
fined tire width. Thinner tires receive the same overall image height but some
areas remain white. This approach can be thought of as something similar to
zero padding to keep the scale between images consistent but simultaneously
ensure that each image has the same dimension. Also, for standard CNNs the
high resolution of the extracted tire circumference image needs to be down-
sampled. Simple downsampling loses a lot of image information. Many pixels
are combined to one, resulting in grooves being closed and therefore lost in
the process. This, combined with the need for more images to train the CNN,
makes other approaches necessary.
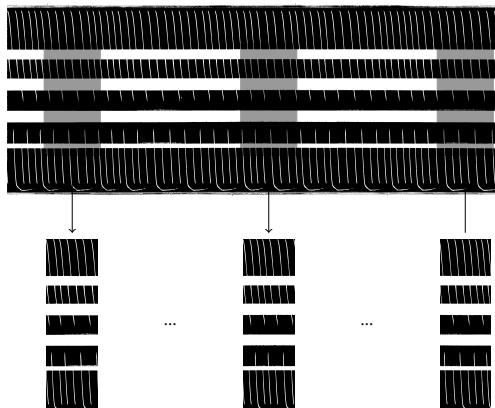
Figure 3.7: Illustration of tire fragmentation process to create multiple tire images from each tire
for usage in CNNs.

This is why one tire image is fragmented into smaller parts by moving a mask
in circumferential direction over the whole structure and saving all masked
images as new tire fragments. Fig. 3.7 visualizes this process. This is assumed
to be a reasonable approach since it represents the actual process of the tire
rolling on the road. The tire structure being in contact with the road changes
due to the randomization of the tire. Nevertheless, no measurements per-
formed for this dissertation indicate that this randomization leads to up and
down swelling of the emitted SPLs over one tire rotation. For this reason, it is

not to be expected that the inclusion of only parts of the individual tires in the images will lead to a high loss of information. Especially because features for shoulder randomization as well as offset randomization can be calculated as additional power density spectra, as explained in [Li16, Leu23]. These spectra can be given to the models as additional information, thus making the information loss through only including tire fragments minimal. Section 4.7 presents models which receive this additional input.

One more step for correct image cropping is necessary. This results from the origin of the coordinate system of each 3D tire model, which is not fixed with regard to its position in relation to the tire. Especially, in case of axis x_2 (see fig. 3.2), only the orientation, pointing into the rim, is defined, but not the location of $x_2 = 0$. Taking the maximum and minimum x_2 values included in fig. 3.7 for cropping is not a robust approach since these values are sensitive to single outliers. Therefore, all grid points of one tire are projected into one plane of a cylindrical coordinate system through their radii. For the tire presented in fig. 3.7, this results in fig. 3.8. From this figure, the correct center location is extracted. For this purpose, points of the tire are extracted whose tire height corresponds to the rim radius plus half of the aspect ratio times the tire width plus 5 mm, rounded to one decimal place. Due to the high resolution of the model and non-equidistant net, multiple of these grid points exist. The relevant center of these points is easiest obtained by rounding all grid points to one decimal and then applying a k-means clustering which results in one cluster for each shoulder.
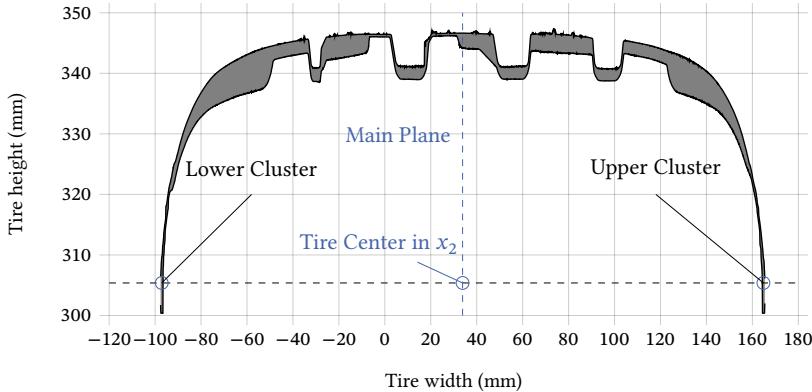
Figure 3.8: Tire silhouette for tire width and center location extraction. X-axis refers to x_2 axis defined in the digitization process of the tire. It is located in the center of the rim, pointing inwards.

By calculating the average x_2 value of these two clusters, the x_2 value of the tire's main plane is obtained. After determining the specific tire center, it is utilized as the center point for the symmetrical cropping process in order to create a rectangular contact patch. The image cropping height is defined as 0.8 times the tire nominal width c_{TW}. The patch length, which corresponds to the image width, is calculated with the assumption that the contact patch area is equivalent for all tires. This leads to

$$l_{\text{contact patch}} = \frac{\frac{g\, m_{\text{vehicle}}}{4}}{p_{\text{inflation}}\, 0.8\, c_{\text{TW}}} \tag{3.3}$$

under the assumption of a rectangle as contact patch. For the used vehicles, a mass of 2200 kg is taken. Even though that calculation includes many simplifications, it can still be expected to lead to satisfying results. This is all the more the case as there are other studies in which the tire contact patch is defined identically for all tires included in these studies [Li17, Moh22]. Since the values obtained for $l_{\text{contact patch}}$ are rather small and might not include enough information for the CNN to learn correctly, each image includes three consecutive contact patches. The mask, which is therefore three times the length of

the contact patch, is iterated over the tire in steps of one tenth of the mask width. This ensures that at least some difference will be recorded between images and also generates more images for the training of the CNN. Furthermore, this approach ensures overlapping between images, which might make it easier for the CNN to learn the relationship between the images. The pressure in the contact patch is set equal to the inflation pressure of the tire during the CB measurements to 2.5 bar. Before using the images as input for the CNN, all cropped images are rescaled to have the same millimeter to pixel ratio of $0.05\,\mathrm{mm}\,\mathrm{pixel}^{-1}$ in image height. The image width is also rescaled to $0.1\,\mathrm{mm}\,\mathrm{pixel}^{-1}$. The scaling in image width is necessary due to the different circumferences of all processed tires. These lead to a different length per degree ratio in each image. The different $\mathrm{mm}\,\mathrm{pixel}^{-1}$ representations are chosen to obtain approximately square tire fragments for tires of average dimensions. After rescaling both dimensions, the images are reframed to a standard size and pasted on a white background of identical size for all tires. The independently appearing maximum image height and width of all tires define the white background image size. The whole cropping, scaling and reframing process leads to tire images as represented in fig. 3.9 of the pixel size 3947 x 4559. In the transformation from images (.png) to PyTorch tensors they are all downscaled to 224 x 224 shape to reduce the CNN dimension even further. As seen in fig. 3.9a the narrowest tire defines the image width while the widest tire, in fig. 3.9c, defines the image height.

(a) Narrowest tire (205)   **(b)** Mean tire (245)   **(c)** Widest tire (285)
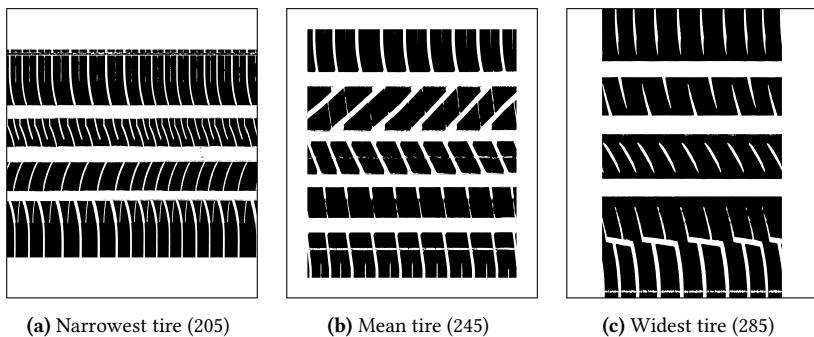
Figure 3.9: Three different tire images after cropping and rescaling to standardized format with pixel size 3947 x 4559. Frames only added for displaying purpose.

4 Development and Evaluation of Statistical and Machine Learning Models

In the following chapter, the data processing as well as modeling processes are explained. The chapter is divided into the different steps of data analysis and modeling. Firstly, some statistical analyses are performed on the data. Secondly, simple benchmark models are presented. Lastly, different ML approaches are presented. These include deterministic models in the sense that once trained, the output for identical inputs will always be identical as well as probabilistic ones. The model complexity is increased over the sections. Model interpretation is conducted with an intuitive data modification as well as through the method of XAI presented in section 2.5. The obtained model performance as well as model interpretation is presented in the same sections as the related model design. In this dissertation, the evaluation of models is mainly based on the RMSE. The RMSE is thereby calculated for the output vector as whole and not for each output considered as individual sample, so that the RMSE of a model with multiple output variables is expected to be higher than the one of models with one output variable.

4.1 Statistical Data Analysis

Before training statistical models in the form of ML or DL, it is necessary to analyze the data on different aspects, such as the general range of input and output values, deviation in repeated measurement, and deviation for both microphone sides. Fig. 4.1 gives an overview of all 1443 conducted measurements

of the 42 different tires, four different measurement vehicles, and other controlled and uncontrolled variables. The displayed SPLs are the ones recorded while the vehicles position is at PP'. For a single nominal speed the maximum spread adds up to the range of 3.96 dB(A) to 6.00 dB(A) for the left and 4.44 dB(A) to 5.85 dB(A) for the right microphone. The considered speeds are in steps of $5\,\mathrm{km\,h^{-1}}$ in the range of $40\,\mathrm{km\,h^{-1}}$ to $90\,\mathrm{km\,h^{-1}}$, all speed measurements are rounded to integers for this calculation.

(a) Left microphone side   **(b)** Right microphone side
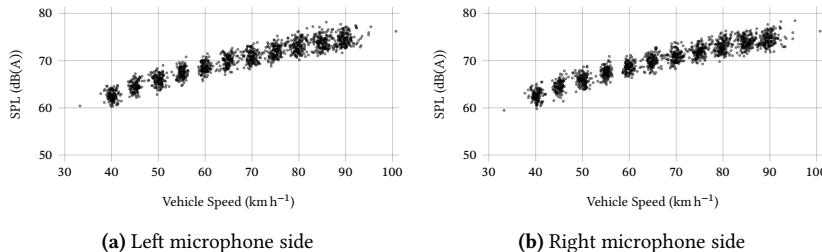
Figure 4.1: Overview of all measurements for both microphones while vehicle is at test track position PP'.

The variability in measurements due to different tires can be approximated through their speed SPL relation in eq. (2.2). This enables the calculation of TRN coefficients a and b for each tire. Since their accuracy has been proven in literature [San02, Leu22], they can be utilized to reconstruct measurements at arbitrary speeds for each tire with high accuracy. Doing so for the whole range in vehicle speeds from $40\,\mathrm{km\,h^{-1}}$ to $90\,\mathrm{km\,h^{-1}}$ leads to the SPLs curves for position PP' for each tire as indicated in fig. 4.2a. The green curves indicate the minimal and maximal reached SPLs over the whole speed range, the blue curve indicates the mean SPL. Fig. 4.2b shows the resulting maximum difference due to different tires at the position PP' for the entire speed range. These are slightly smaller than the recorded values from the measurements. However, this is not surprising, as the displayed values form an average over the left and right microphones and general measurement deviations are also reduced by calculating the TRN coefficients.
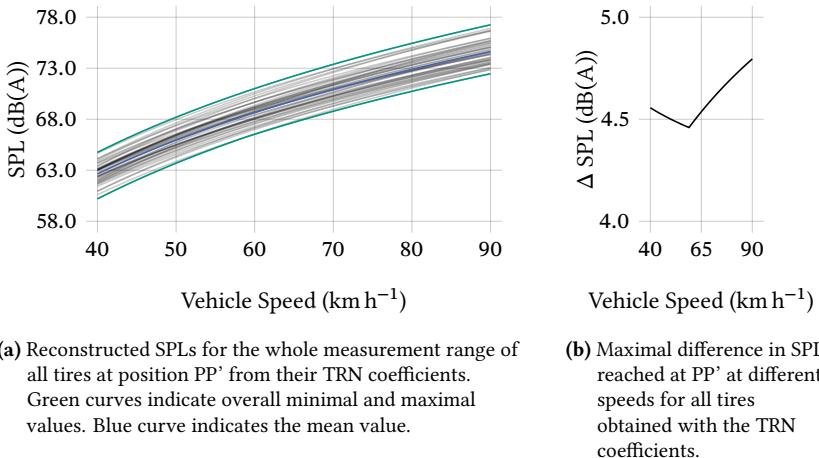
(a) Reconstructed SPLs for the whole measurement range of all tires at position PP' from their TRN coefficients. Green curves indicate overall minimal and maximal values. Blue curve indicates the mean value.

(b) Maximal difference in SPL reached at PP' at different speeds for all tires obtained with the TRN coefficients.

Figure 4.2: Reconstructed SPLs for all tires over the whole speed range at PP' with associated difference in SPL obtained at each speed.

Other than the total spread in measurements especially the difference in recorded SPLs during one run between the left and the right microphone is interesting, since this mainly represents the variability which is assumed to be seen in models with output values for both microphones as well. In case of the measurement deviation between both microphones not only the absolute values but especially the distribution is of interest. Fig. 4.3a shows the recorded deviations. The mean recorded deviation is calculated to $-0.04\,\mathrm{dB(A)}$.

The distribution of the deviation can be approximated through a Kernel Density Estimation (KDE). Here, a Gaussian kernel with a bandwidth of 0.2 is applied on the data. Fig. 4.3b shows the resulting Probability Density Functions (PDFs) for different vehicle positions in the range of $-10\,\mathrm{m}$ to $10\,\mathrm{m}$ in steps of $2\,\mathrm{m}$. The PDFs do not change substantially. Also, they are all centered around the value of zero deviation. Combined with the Gaussian shapes of the PDFs this indicates, that the overall measurement differences between both microphones can be assumed to be Gaussian distributed with zero mean.
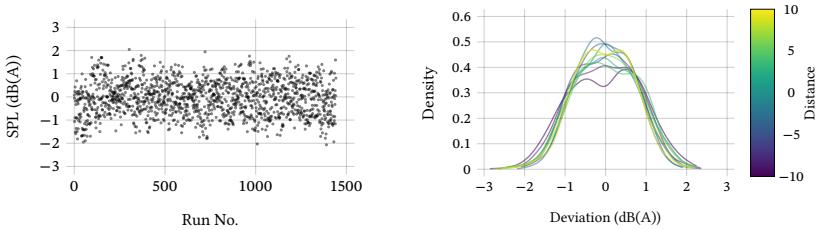
(a) Deviation of left and right microphones for all runs at position PP'. Mean value at $-0.04\,\mathrm{dB(A)}$.

(b) PDFs for SPL deviation between left and right microphone for different vehicle positions on the measurement track.

Figure 4.3: Overview of all measurements for both microphones with deviation between microphones in absolute values and as probability distribution.

Another feature for which the impact can be compared to many investigations from literature is the temperature impact on the emitted SPL. According to UNECE regulation no. 51 [UNE18] the air temperature range during measurements is supposed to be in the range of 5 °C to 40 °C. UNECE regulation no. 117 [UNE22] demands additional road surface measurements which must be in the range of 5 °C to 50 °C. Regulation 117 also implements a temperature correction of the SPL, as already presented in eq. (2.4).

Engineers at Mercedes-Benz (MB) however designed a slightly different temperature correction curve based on their many years of experience. This correction is defined as

$$L_P(T_{\mathrm{ref}}) = L_P(T_{\mathrm{air}}) + \tau_{\mathrm{MB}} \log\left(\frac{T_{\mathrm{air}}}{T_{\mathrm{ref}}}\right) \tag{4.1}$$

with τ_{MB} being $2.60\,\mathrm{dB(A)}$. This temperature correction was investigated in an earlier publication [Leu22]. The analysis concluded that the MB temperature correction leads to overall higher CoDs of tire models than the correction from the standard. This is why the MB correction is considered to be a better representation of the actual temperature relation with the SPL than the correction from the standard. This is not surprising since its steadiness alone is physically more plausible. [Leu22] Fig. 4.4 shows a comparison of the adjustment of the SPL resulting from the two different corrections.
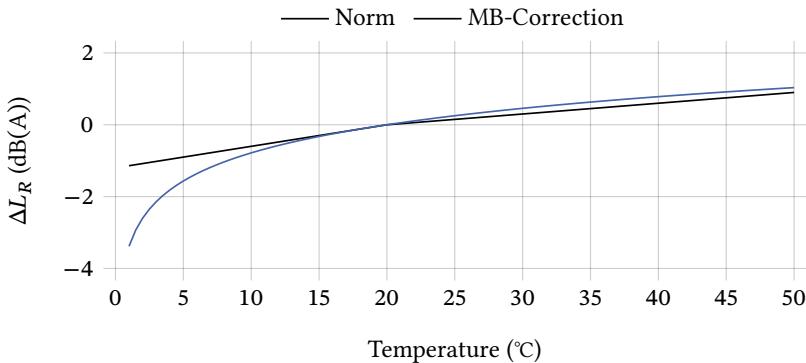
Figure 4.4: Two different temperature correction curves for the absolute SPL. Reference temperature is 20℃ as indicated by the zero-crossing of both curves at 20℃. The illustrated curves show the correction for the air temperature measured on the test site. [Leu22]

The air temperature of the measurements ranges from 13.40 °C to 31.46 °C and the road temperature from 15.83 °C to 49.37 °C. It can be noted that the temperature range in which measurements are performed makes it almost trivial which correction curve is applied since no large difference occurs. Table 4.1 presents the overall deviations due to temperature for both temperatures and correction curves.

Table 4.1: Absolute temperature influence on the SPLs of the entire data set, calculated with two different correction functions and two different temperature measurements.

Temperature	Norm		MB	
	Min	Max	Min	Max
Air	−0.40	0.34	−0.45	0.51
Road	−0.25	0.88	−0.26	1.02

In literature some different opinions on which temperature results in the most accurate adaptation of the SPL exist. It is also questionable if the correction curves need to be adapted for different temperatures. As indicated in fig. 4.5, a linear relation between air and road temperature of all performed measurements can be found. This relation reaches a CoD of 0.69 for

$T_{\text{road}} = 1.63T_{\text{air}} - 2.34\,°C$. More interesting than the linear relationship is the actual range of the recorded values, which is almost twice as large for the road temperature as for the air temperature. While all the recorded air temperatures are distributed around the reference value of $23\,°C$, most road temperatures are located around $37\,°C$ as indicated by the PDFs in fig. 4.5. The PDFs are estimated with Gaussian kernels and a bandwidth of two. The higher overall values in road temperature are not surprising since all measurements are conducted outdoors, which makes it nearly impossible to prevent the surface from warming up due to solar radiation. This also leads to much higher local fluctuations due to measurement location as well as temporary fluctuations due to e.g., cloud fronts for the road temperature compared to air temperature. This observation aligns with the literature reported on in the temperature paragraph in section 2.1.2.
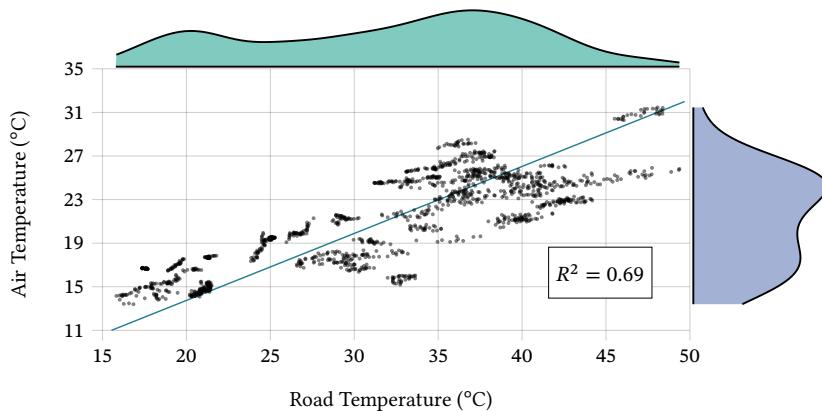
Figure 4.5: Relation between air temperature and road temperature measurements for all runs, at position PP' with PDFs and regression curve calculated with least-squares.

In the following sections, tire parameters are considered as features in ML models and the models are used to explain the impact of single features on the recorded SPL through methods of XAI. For features with high covariance, this can lead to fluctuating feature importance since the model somehow randomly assigns the weights to highly correlated features. Therefore, fig. 4.6 shows the

covariance between features calculated for the overall amount of 145743 data points. The tire nominal width, tire aspect ratio, and the rim diameter have rather high correlation coefficients. In particular, the correlation between the tire nominal width and the rim diameter is not surprising. This can be explained with the overall industrial trend of growing diameters and widths due to design reasons, as well as other changing engineering requirements like heavier vehicles that demand higher possible loads as well as growing engine power which results in larger brakes and therefore demand for more space. The negative correlation between the tire nominal width and the aspect ratio is also explained due to the definition of the aspect ratio, which defines the sidewall height in percent of the nominal width. Meaning not reducing the aspect ratio with increasing width would lead to exceptionally high tire sidewalls. The same principle applies for the rim diameter and the aspect ratio. With growing rim diameter, the aspect ratio generally decreases. This leads to a compensation of the outer tire diameter. The speed index only shows somehow higher correlation with the aspect ratio. This can be explained through the construction of performance tires, which tend to have large width and low aspect ratio to increase transversal stiffness. At the same time these tires are designed for high speeds and thus have high speed indices. The load index only shows moderate correlation to other parameters. In appendix B more correlation analysis can be seen.
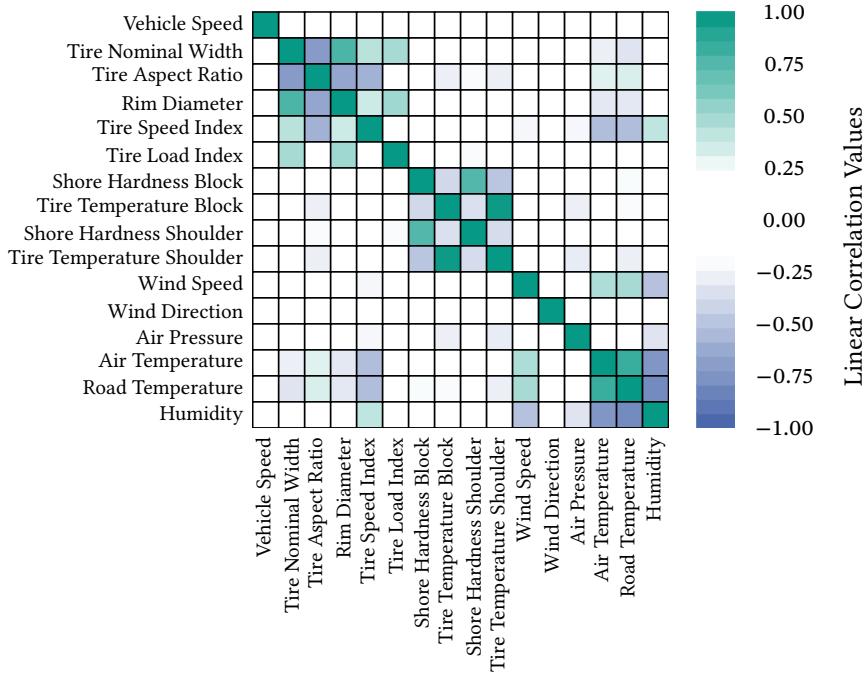
Figure 4.6: Feature covariance as heat map for input features with each point on test track as individual data point.

4.2 Linear Regression Models

Before training complex ML models, it is reasonable to investigate some simpler models. These can be used as benchmark models for more complex models to evaluate the trade-off between model performance and complexity. These benchmark models mostly do not require complex optimization strategies and allow for easier model interpretation.

The numerical data recorded during the measurements can be used for linear regression models. Three different modeling approaches are investigated.

4.2.1 TRN Coefficients

As described in section 4.1, the TRN coefficients are calculated from eq. (2.2) for each tire. These parameters can be averaged to create a predictive model. To ensure comparable conditions for the accuracy, similar to later evaluated ML models, a train test split with LOOCV is applied. The omitted tire is therefore predicted using the average coefficient values and the deviations from the actual measurements are calculated. Since this model does not require any knowledge of further tires for the prediction, it can be interpreted as the most simple possible benchmark model.

The data sets are not specially filtered, leading to slightly unbalanced sets since the number of runs per tire is not exactly the same. However, this is considered the best approach when it comes to model comparison since the same runs are used furthermore for the ML models. The decision to not restrictively filter the data is taken to ensure that the spread in variables as the air temperature or wind speed is as high as possible. Also, as shown in a previously published research paper, the impact of slightly unbalanced data sets for multiple linear regression models is not essential [Leu22].

The TRN coefficients are calculated as average values for the left and right sides. Fig. 4.7 displays the model input and output. This sketch will reoccur throughout the whole methodological part of this dissertation to allow for a quick overview of the model design.
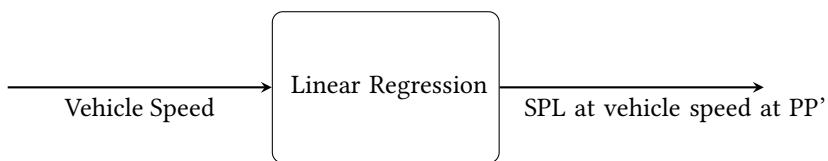
Figure 4.7: Principle sketch of TRN coefficients model.

The LOOCV with the RMSE of the predictions is used to evaluate the prediction quality. The same evaluation is used later for the ML models. The error

vector necessary for the calculation of the RMSE is constructed by subtracting the predicted output from measurements for the left as well as the right microphone.

This approach leads to the results in table 4.2. These results, as well as all further results, are however dependent on the chosen set of tires and do not guarantee to represent the overall market.

Table 4.2: Performance metric for TRN coefficients model calculated with LOOCV. All values expect Mean_{R2} in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

$\text{Mean}_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	Mean_{R2}
1.56	0.78	3.60	0.60	0.89

To gain more insight into the loss, fig. 4.8 displays the correlation between the coefficients a and b as well as their PDFs. For the calculation of the TRN in this dissertation, the reference speed v_{ref} is set to $1\,\text{km}\,\text{h}^{-1}$. This implies the coefficients, especially coefficient a, should not be interpreted as SPL at the reference speed but they are merely mathematical coefficients used in the equation. It is visible that a high correlation between both tire coefficients exists. A linear regression model of both values, as indicated in the figure reaches $R^2 = 0.87$ with a mean residual of zero (-5e-15) and a Standard Deviation (STD) of residuals of 0.61. Considering these values as well as the fact, that the 95 % confidence interval for the a values lies in the range from 7.25 dB(A) to 9.02 dB(A) and for b in the range from 33.51 dB(A) to 34.55 dB(A), it is obvious that even the least informative model is not completely off in the average prediction. However, especially tires which are far away from the mean values, lead to poor prediction accuracy in the model. This leads to the maximum RMSE of 3.60 dB(A). Furthermore, due to its simplicity, this benchmark model cannot provide any further knowledge on feature importance of the tire and is not considered useful for predictions of new tires since it only predicts the mean values of the speed relation in the training data. However, this is a direct result of the input values which are used in this model.
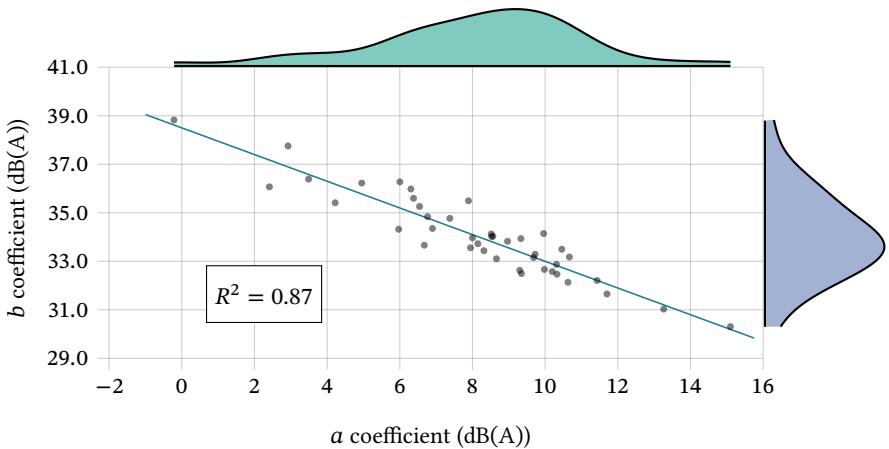
Figure 4.8: Analysis of correlation of TRN coefficients with their independent PDFs. As reference speed v_{ref} is defined as $1\,\text{km}\,\text{h}^{-1}$.

4.2.2 Linear Regression

As second benchmark model, a linear regression with more numerical input data is considered. No interactions between variables are considered and only the vehicle speed is preprocessed to be $\log_{10}\dfrac{v}{v_{\text{ref}}}$ with $v_{\text{ref}} = 1\,\text{km}\,\text{h}^{-1}$. This is applied because the literature provides evidence of this connection ([Nil76, San02, Leu22]) and disregarding this can be considered to intentionally worsen the model accuracy.

As stated above, the second benchmark model takes more values as input. These partly describe the tire itself, the environmental conditions, and the operation conditions. Fig. 4.9 shows a full principle sketch with the corresponding input and output values of the model.
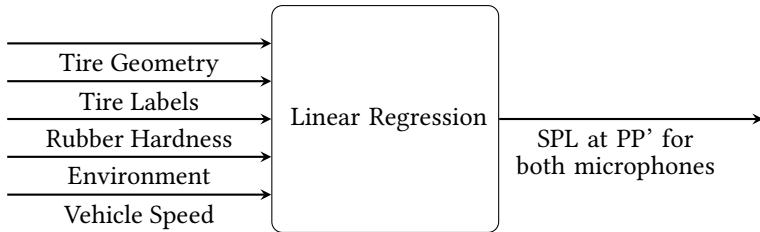
Figure 4.9: Principle sketch of the linear regression model. Rubber hardness incorporates two hardness values with correlated surface temperature. Environment incorporates the wind speed and direction, air and road temperature, air pressure, and humidity. Tire geometry incorporates the tire nominal width, the aspect ratio, rim diameter, and profile depth. Tire labels refer to the load and speed index.

The performance metric for the linear regression model is displayed in table 4.3. The model outperforms the first presented model which only used the TRN coefficients in the maximum and the mean RMSE.

Table 4.3: Performance metric for linear regression calculated with LOOCV. All values expect $Mean_{R2}$ in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	$Mean_{RMSE}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$Mean_{R2}$
Linear Regression	1.54	0.65	3.04	0.68	0.90
TRN Coefficients	1.56	0.78	3.60	0.60	0.89

The mean RMSE with 1.54 dB(A) is rather good for such a simple linear regression model, however a maximum RMSE of 3.04 dB(A) is very high and shows that the model does not generalize. One large advantage of the simple regression models compared to the latter build machine learning models is the easy interpretability. Even though the model does not result in a low RMSE, the associated CoD is high, indicating that the model does capture a high percentage of the data variance correctly. It is therefore still interesting to look into the model reasoning.

The importance of feature x_i can be analyzed through

$$(\max(x_i) - \min(x_i))\, mean(\beta_i) \tag{4.2}$$

which calculates the absolute difference in SPL caused by one feature over its recorded domain on basis of the mean correlated regression coefficient β_i of all different models obtained through the cross validation as average for the left and right microphone.
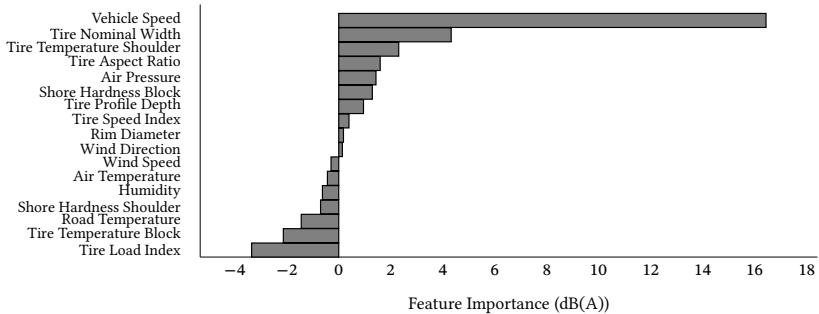
Figure 4.10: Linear regression model feature impact, associated p-values of the regression coefficients are presented in table A.6.

As indicated in the mean feature importances in fig. 4.10, the major variability in the SPL results from the vehicle speed, tire nominal width, and tire load index. The overall high values especially in the speed importance is as expected from earlier investigations where only the speed was considered. The low absolute importance of the wind speed and wind direction seem plausible since measurements are conducted to a maximum wind speed of $5\,\mathrm{m\,s^{-1}}$ which is assumed to not impact the SPLs gravely. Still, the observation that with higher wind speeds, the SPL decreases is not plausible, the irrelevance of both variables is also indicated by their p-values which are calculated to 0.38 and 0.60 for wind speed and wind direction. The importance of the shore values lacks explainability. While the shore values in the shoulder region correlate negatively with the SPL, the shore values in the tread region correlate positively with the SPL. This can be a result of the position at which the shore hardness measurements are conducted. The shoulder measurement is under normal conditions not part of the contact patch, so that the relevant indication of the model is an increasing SPL for increasing rubber hardness. Generally,

it is also questionable if the model weights the impact of the shore hardness at the shoulder well, since its p-value reaches 0.25. The associated temperatures show opposing behavior. In case of the air and road temperature, the model captures the overall understanding of a negative correlation between the temperatures and the SPL in the range from 5 °C to 50 °C. The effect is only spread over the two variables passed independently to the model. Still, the p-value of the air temperature is greater than the one of the road temperature which might indicate that the road temperature is generally the more important variable. Lastly, the figure shows that a strong negative correlation between the load index and the SPL exists. This could be a result of higher tire mass and stiffer inner structure which generally come along with higher load index and lead to less vibration and therefore less noisy tires. However, it can be possible that the trends in fig. 4.10 are impacted by the high multicollinearities of some variables. All p-values are available in table A.6.

The feature importance can be examined with a sanity check. This is possible through the well-known and proven speed relation of the TRN. Combining this with the observation of the distribution of the TRN coefficients leads to the approximations

$$I_v \approx \overline{b} \log\left(\frac{v_{\max}}{v_{\min}}\right) \tag{4.3}$$

$$\sum_{i \in F \backslash v} I_i \approx \overline{a} \quad . \tag{4.4}$$

where: $\overline{a}$ Arithmetic mean of TRN coefficients a
         $\overline{b}$ Arithmetic mean of TRN coefficients b
         F Set of all features
         I_i Feature importance of feature i
         I_v Feature importance of the vehicle speed i
         $v_{\min}$ Minimum speed in data set
         $v_{\max}$ Maximum speed in data set

In case of eq. (4.3) the values are similar with 16.43 for the left hand side and 16.38 for the right hand side, which indicates that the model correctly reproduced the impact of the vehicle speed on the SPL. In case of eq. (4.4) the

values are with 3.64 and 8.13 far more apart. This is not entirely surprising since the linear regression model is also off by a mean RMSE of 1.54 dB(A), therefore the sanity check can be considered working.

4.2.3 Lasso Regression

The third benchmark model is a lasso regression, which has, as shown in fig. 4.11, the same input variables as the linear regression model.
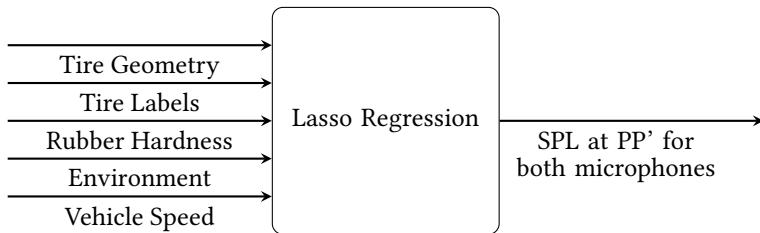

Figure 4.11: Principle sketch of the lasso regression model. Rubber hardness incorporates two hardness values with correlated surface temperature. Environment incorporates the wind speed and direction, air and road temperature, air pressure, and humidity. Tire geometry incorporates the tire width, the aspect ratio, rim diameter, and profile depth. Tire labels refer to the load and speed index.

The lasso approach (explained in section 2.4.1) applies an L1 norm regularization on the regression coefficients to avoid large coefficients. This process makes use of a hyperparameter λ, that scales the impact of the L1 norm. In case of the lasso regression, a hyperparameter optimization of λ can be performed. The hyperparameter for the present problem is evaluated through a grid search with values in from 0.01 to 1 with steps of 0.01. For each data split the best hyperparameter is obtained by evaluating the train performance through 5-fold-cross-validation and comparing the obtained training performance to those with other hyperparameter values of the search space. The optimization and calculation is done with the Python package Scikit-learn [Ped11]. In all 42 data splits, the λ values leading to the lowest RMSE are determined from 0.02 to 0.04. As seen in the associated performance metric in

table 4.4 this leads to a 0.09 dB(A) lower mean RMSE and 0.09 dB(A) lower maximum RMSE compared to the linear regression model.

Table 4.4: Performance metric for lasso regression calculated with LOOCV. All values expect Mean$_{R^2}$ in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	Mean$_{\text{RMSE}}$	STD$_{\text{RMSE}}$	Max_{RMSE}	Min_{RMSE}	Mean$_{R^2}$
Lasso Regression	1.45	0.55	2.95	0.70	0.91
Linear Regression	1.54	0.65	3.04	0.68	0.90
TRN Coefficients	1.56	0.78	3.60	0.60	0.89

Fig. 4.12 shows the feature importance for the linear regression model in fig. 4.12a and for the lasso model in fig. 4.12b. For easier visual comparison, fig. 4.12b presents them sorted according to their importance in the linear regression model. It seems that the overall trend of feature importance remains similar. However, it is especially notable that the temperatures of the shore hardness measurements lose substantially in impact. This is what would generally be expected. Summarized, the lasso regression almost suppresses nine features completely. This exact behavior is why lasso models are sometimes considered as models for feature selection. The coefficient vector represents a sparse solution and therefore exactly what is expected from a lasso regression. Interestingly, the impact of the environmental air pressure remains intact and also reaches high significance as indicated by their p-values (table A.6). This can be explained by the slight imbalance in the data, generally more tires are measures at higher air pressures than lower air pressures. These furthermore show slightly higher SPLs than the ones measures at lower air pressures which is also a result of the observation that slightly more data points at higher speeds exist at higher air pressures than at lower air pressures.
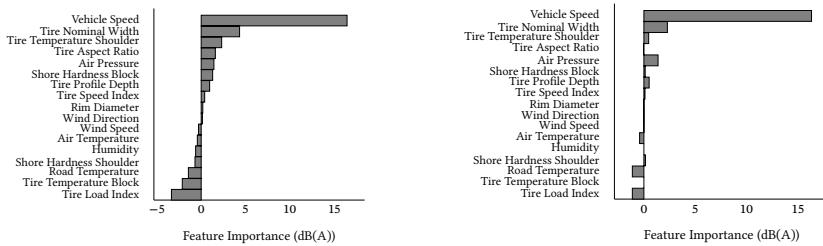
(a) Linear regression model feature impact

(b) Lasso regression model feature impact

Figure 4.12: Comparison in feature importances obtained for the linear and lasso model, sorted according to the feature importance obtained for the linear regression model. Both figures display the feature importance as average of the LOOCV model splits for the average of the left and right microphone.

4.3 Artificial Neural Network

The following section is divided into two subsections. First, an overview of the scope of the used ANNs is given and a hyperparameter optimization is performed. The optimized models are then used to approximate the generalization performance.

4.3.1 Model Definition and Hyperparameter Optimization

All models designed so far are multiple linear regression models, which cannot depict nonlinearities between features but only in variables themselves. Even though they would also allow for higher orders of variables in features, this was not taken into account. However, the relation between tire parameters and their sound emission is highly nonlinear. For this, ANNs as described in section 2.4.2 are useful. All input parameters used previously in this research are solely numeric values so that they can be directly used as input for an ANN. The ANN also has the advantage that the microphones on the left and right sides of the test track can be used in a multi-output regression. This allows for an interdependent calculation of the weights. The model outputs for both microphones are thus related, which helps in avoiding unsteady solutions. In

the earlier presented linear models, both sides are completely independent. This can in essence lead to predictions that diverge largely for both sides.

Additionally to the presented numeric values used in the linear model approaches, the ANN receives information about the measurement vehicle, the tire manufacturer, and the driving direction. Fig. 4.13 shows the model sketch. The driving direction indicates from which side the symmetric test track is entered.
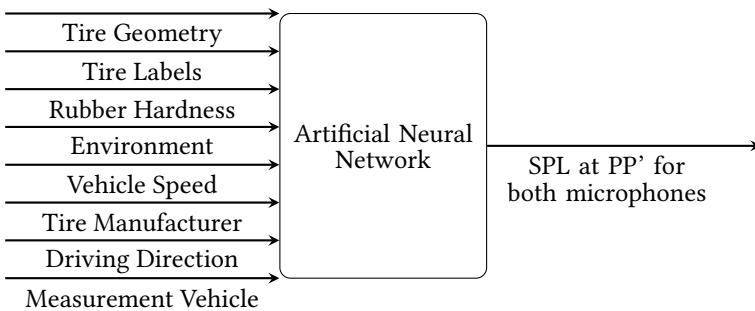
Figure 4.13: Principle sketch of the ANN. Rubber hardness incorporates two hardness values with correlated surface temperature. Environment incorporates the wind speed and direction, air and road temperature, air pressure, and humidity. Tire geometry incorporates the tire nominal width, the aspect ratio, rim diameter, and profile depth. Tire labels refer to the load and speed index.

The measurement vehicle and tire manufacturer cannot be directly included as input in the ANN. This is due to them being on nominal scale (e.g., words in case of the manufacturer) and series numbers (in case of the vehicles). While the second is a numeric format and therefore generally usable in ANNs, giving these numbers directly to the ANN is not expected to result in realistic outputs. This is because the numbers defined by the vehicle manufacturer do not have a strict mathematical relation.

These issues can be overcome through applying encodings to the input. Section 3.3 gives an explanation on the one-hot encoding. The one-hot encoding of the measurement vehicle, the tire manufacturer, and the driving direction

leads to them being in linear numeric space. As such, it is more reasonable to use them as input in the ANN.

In case of the four different measurement vehicles this leads to the transformation as shown below.

$$
\begin{array}{c}
\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\
\begin{array}{l}
167 \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \\
212 \longrightarrow \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \\
243 \longrightarrow \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \\
293 \longrightarrow \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
\end{array}
$$

The input vector representing the vehicle labeled as 212 (E-Class) is represented as the second row of the matrix, so that $x = (x_1, x_2, x_3, x_4)$.

The measurement vehicles could also be described with deeper features which include actual relations and scalability. These could include the vehicle mass, length, engine power, drag coefficient, wheelhouse geometry, or ground clearance. However, in an earlier publication it was found, that the impact of five different measurement vehicles on the TRN under measurement conditions according to UNECE regulation no. 51.03 and UN regulation no. 117 are negligible within the group of investigated vehicles [Leu22]. Dissolving the vehicles in deeper features is therefore not expected to lead to higher model accuracy. In cases where more than four different vehicles are used for measurements, deeper features could be useful for the model, however in this case due to low variance it is not expected to produce better models. Still it is reasonable to allow for a distinction of the different vehicles which is achieved through the one-hot encoded values as input.

The one-hot encoded values can be used as single neurons for each value x_i of x in the input layer. Guo et al. show that it might be better to include the one-hot encoded variables in an additional input layer, the so called embedding layer [Guo16]. This approach breaks down the one-hot encoding to a predefined number of neurons. The further used ANN includes three embedding structures: the first one receives the fourteen different tire manufacturers as

input, the second one receives the four different measurement vehicles as input, and the third one receives the shore hardness on shoulder and block with the corresponding tire surface temperature as input. Each of these structures breaks down their inputs to one output. This definition of the terminology clarifies, that the embedding happens in one layer, yet the single values that are embedded are completely independent. Therefore these embeddings are referred to as embedding structures. All three embedding structures break down their input to one neuron so that the core ANN receives a total number of 17 features as input. The choice to apply embedding on the shore hardness is taken due to its interdependence between both hardness values as well as the associated temperatures, as seen in fig. 4.6.

The equation of how the encoded values are combined is trained during network training and can also include nonlinearities through activation functions. Fig. 4.14 shows the structure developed for hyperparameter optimization.
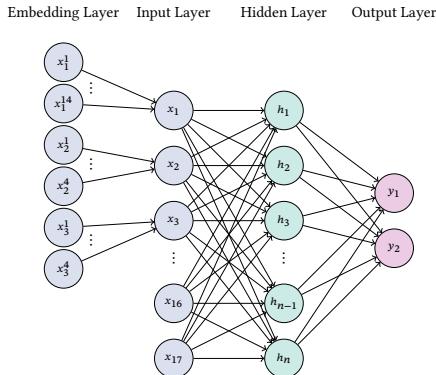
Figure 4.14: Principle sketch of ANN architecture with embedding layer including three embedding structures.

A Python framework for hyperparameter optimization is Optuna [Aki19]. In cases of multi-objective Optuna uses the *Non-dominated Sorting Genetic Algorithm II* (NSGA-II) [Deb02] sampler. The Optuna package also allows ranking

the impact of different hyperparameters on the models performance through an adaption of the Analysis of Variance (ANOVA) [Hut19].

During the hyperparameter optimization the following parameters are investigated:

- Activation function in embedding layer: Rectified Linear Unit (ReLU), no activation
- Activation function in hidden layers: Sigmoid, Tangens Hyperbolicus (Tanh), ReLU, and Leaky Rectified Linear Unit (LeakyReLU) with negative slope of 0.1
- Activation function in output layer: ReLU, no activation
- Number of layers: embedding layer, input layer, 1-4 hidden layers, output layer
- Batch size: 4, 16, 64
- Learning rate: 1e-4, 1e-3, 1e-2
- Dropout rate: 0, 0.1, 0.2, 0.4

The dropout layer is located after the first hidden layer. The dropout layer deactivates a random number of neurons independently in each forward pass. Whether a neuron is deactivated is determined by a Bernoulli distribution for the given probability. The number of neurons in each layer is indirectly defined by the number of layers. A subsequent layer has four times or a quarter of the number of neurons of its preliminary layer, depending on whether its layer count (counted from zero) is smaller or larger/equal the number of layers divided by two. Model training is performed for a maximum of 5000 epochs, but early stopping applies if for 100 epochs the test set performance does not improve. Overall, 200 hyperparameter combinations (further referred to as trials) are evaluated. The Python package Optuna handles the parameter selection for each model. The optimization process is defined as multi-objective. The training and test loss, which are both minimization criteria, are the objectives of the optimization. The decision to optimize both, the training and

test loss, is made to reach a reasonable compromise between potentially over- and underfitting.

Hyperparameter optimization is performed on three different data splits to ensure that the ANNs architecture is not specifically optimized for one tire. For this, tires 0, 20, and 40 from table 3.1 are chosen as test tires and the data sets are constructed according to LOOCV. Each of these hyperparameter optimizations is conducted completely independent for 200 trials. Since the model parameter settings are defined during the hyperparameter optimization, they also vary for the different optimizations. The optimization uses Adam as the optimizer, and the loss function is MSE. All further presented ANNs and CNNs are built in PyTorch [Pas19].

The input data which is not one-hot encoded is standardized with the z-score normalization presented in eq. (3.2). The transformation is applied on each feature individually. This is done since literature shows that feature scaling speeds up the learning process through weighting all attributes equally [Han11].

Fig. 4.15 shows the three different hyperparameter optimizations with importance of each hyperparameter for the test loss. The strongest agreement seems to lie in the importance and the activation function after the output layer, whereby most of the values of the theoretical Pareto front show that it is better not to use an activation function. However, solely considering the theoretical Pareto front of the study might be misleading. Fig. 4.16 shows all 200 trials with test and train loss. The train loss is thereby always calculated for the model which performed the best on the test data. Assuming a model to neither under- nor overfit, the multi-objective optimization with train and test loss consists of two linearly dependent objectives. Furthermore, the augmentation shows that many more trials exist which are almost not differentiable in performance from the ones defined by the theoretical Pareto front. Retrospectively, it is reasonable to argue that the data does not exhibit an explicit Pareto front since no real trade-off curve between test and train loss is seen in fig. 4.16.
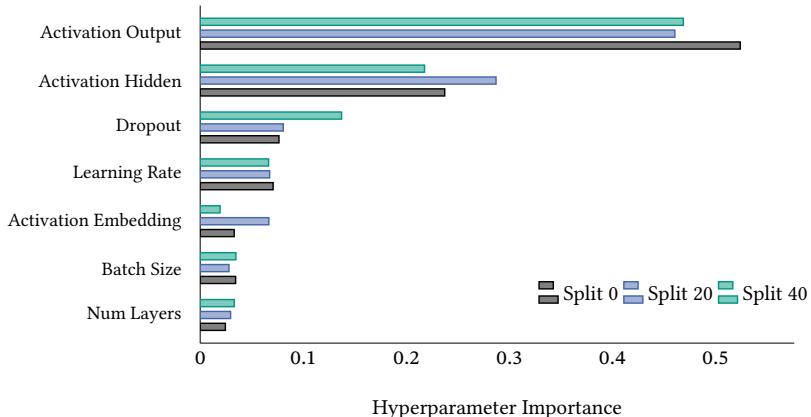
Figure 4.15: Hyperparameter importances for test data set for optimization with three data splits for a study including 200 different parameter settings for the ANN.
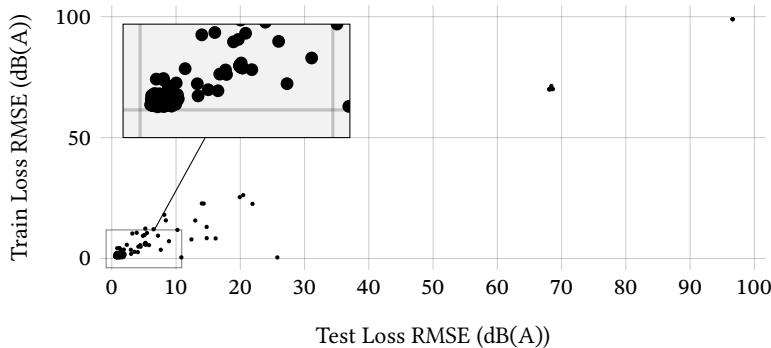
Figure 4.16: Train over test losses for hyperparameter study of first data split in ANN.

The most important hyperparameters are the activation functions in the output and the hidden layers. Considering each 10 best performing models on the test sets defines LeakyReLU 14 times as the most promising activation function for the hidden layers. The importance is rather consistent for test and train data set, which is shown in table A.7.

Taking a closer look into the optimization history, which shows the minimum test loss for all 600 trained models, in fig. 4.17, it seems that the study is mainly impacted by large outliers. The associated distribution is shown in fig. 4.18 and demonstrates that 6 models reach losses higher than 90 dB(A) and 36 models are in the range from 65 dB(A) to 70 dB(A) . The analysis of large outliers shows that in these cases, one of both microphone values is constantly predicted as zero if the last layer is ReLU activated, otherwise they even reach negative values, while the other side predicts reasonable values. These are in the range of 60 dB(A) as RMSE. Models which predict zero for both sides have around 90 dB(A). In case of smaller outliers in the range of 10 dB(A) to 20 dB(A) RMSE, the predicted values for left and right side seem to be consistent but off by a certain amount. The models in the RMSE range of five seem to predict almost the same SPLs for any input, which results form large bias terms. This is observed to result from the sparsity of data in combination with the activation function in the hidden layers. These models mostly use Tanh or Sigmoid as hidden layer activation which reach saturation for some input values. That input value in these cases is smaller than the data grid, which leads to constant predictions.
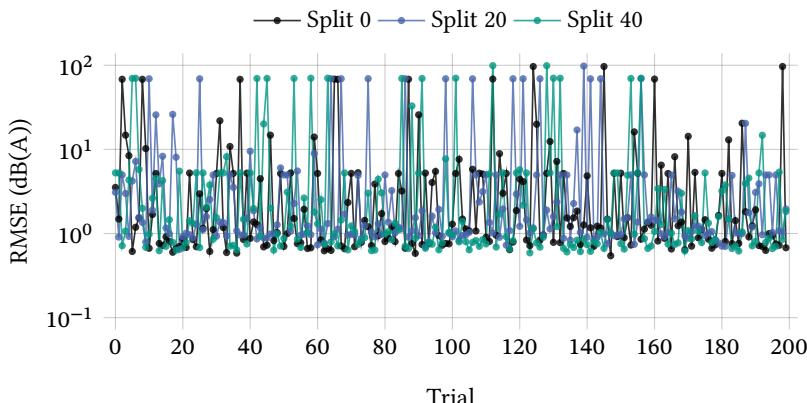
Figure 4.17: Optimization history for three different hyperparameter studies of an ANN. Each study uses a different split for test loss estimation. The figure displays test losses only.
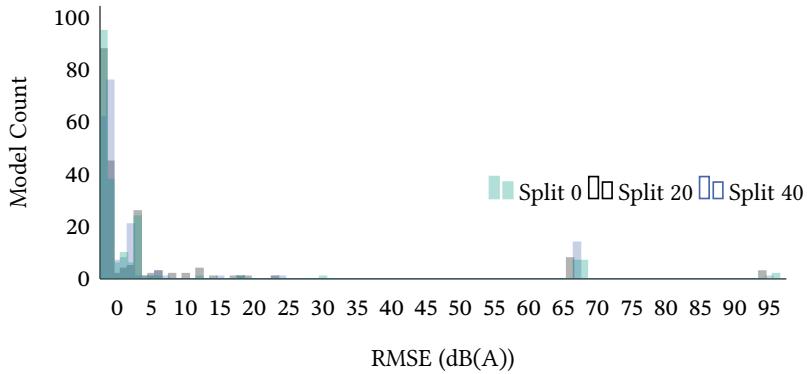
Figure 4.18: Histogram for three different hyperparameter studies of an ANN. Figure displays test losses only.

All together, fig. 4.17 and fig. 4.18 allow two conclusions. First, the majority of models (404/600) will reach acceptable accuracy (RMSE below 2.0 dB(A)) on the test set, no matter the choice of the hyperparameter settings, at least in the investigated ranges of hyperparameters. Second, no optimal hyperparameter settings for the three test studies exist.

4.3.2 Performance Evaluation

Since the hyperparameter study did not result in distinct hyperparameter setting for the investigated models for the LOOCV investigation the following settings are chose as a compromise:

- Activation function in hidden layers: LeakyReLU

- Activation function in output layer: ReLU

- No activation function in embedding layer

- Number of hidden layers: 4

- Batch size: 4

- Learning rate: adaptive between 1e-5 and 1e-2

- Dropout rate: 0.1

Throughout the hyperparameter studies, the performance of the model is assessed by examining the combination of test and train loss. This is decided to build a compromise between under- and overfitting, but it does not guarantee that the best performing model for the test set is chosen. To avoid this behavior, in the LOOCV for each split two models are kept and investigated on their losses. The first model is the model that reaches the minimal test error, not taking the training loss into account, these models are further indexed with **Test optimized**. The second model represents the model with minimal sum of test and training loss. In the following section the second model is referred to as the Test-Train (TT) model performance, these models are for visualization purposes not indexed specially. After each epoch the whole training set is passed through the model in training to calculate the exact loss on the training set equivalently to the test loss. The hyperparameter study furthermore showed that rather high learning rates are necessary for the model to not get stuck in local minima during training. These learning rates tend to lead to strong oscillation in the near region of the expected global minima. This observation leads to the implementation of an adaptive learning rate. This principle was introduced by Smith [Smi15]. In this dissertation, the cyclic range is defined in the range from 1e-5 to 0.01, the learning rate is re-scaled each cycle with an exponential function as seen in fig. 4.19. One cycle represents 10 steps upwards and 50 steps downwards. After each epoch, one step in the learning rate is taken. The model training includes 600 epochs. Early stopping, as in the hyperparameter study, is not applied since this could lead to interruption in the training before reaching small learning rates.
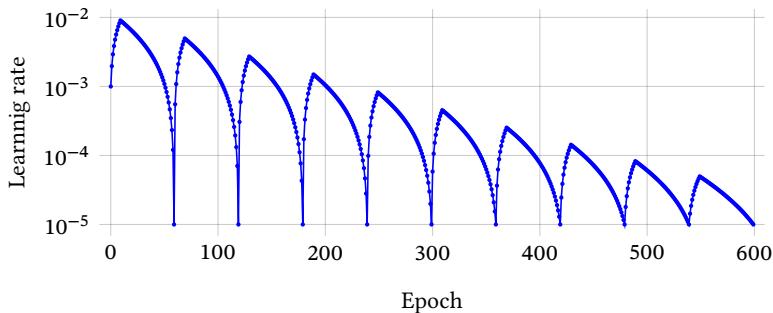
Figure 4.19: Cycling learning rate for the ANNs.

The resulting RMSE of each data split is shown in fig. 4.20 for the models with combined train and test loss as evaluation criteria. The diagram shows that for the majority of models the test performance reaches less than 0.3 dB(A) higher RMSE values than on the train data set. The fact that some models score up to 0.4 dB(A) higher RMSE values on the train set than the test set is assumed to be caused by the method of choosing the best model through the combined test and train loss. The changes in test loss for each data split are caused by the imbalance in test and train data. Through the complete elimination of one tire in the training set and using this data as test set, the data sets are not balanced at all and therefore lead to higher fluctuation than k-fold methods that ensure equal populations in test and train set. Also, some tires have larger deviations in their true data between single runs or between the microphones, which naturally causes their losses to be larger. Since the test set is only constructed with one tire, these effects dominate the changes in test performance. However, it is visible that the ANN approach performs better than the linear and lasso regression models on all shown metrics. It is also visible from fig. 4.20 that the variance in the performance on the training set is smaller than the variance of the test data set. This can be interpreted as an indicator that the data split does not impact the training process hugely. Training performances that vary more than test performances would have to be examined since the training data set does only change by approximately 1/42 for all different splits so that no large changes in the training performance should appear. The corresponding performance metric is shown in table 4.5.

All values $Mean_{\text{RMSE}}$, STD_{RMSE}, and Max_{RMSE} are herein metrics for the performance on the test data set. Only $Mean_{\text{RMSE, train}}$ describes the performance of the models on the training data set. This nomenclature is used in all following performance metrics.
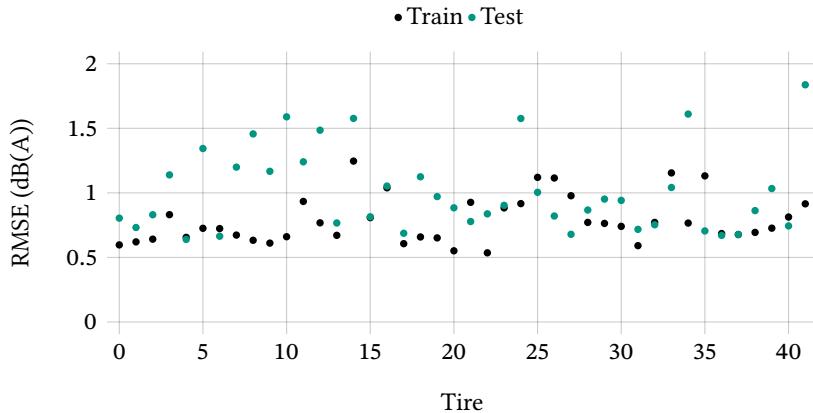
Figure 4.20: RMSE for LOOCV data splits for ANNs defined as Test-Train (TT) models.

Table 4.5: Performance metric for ANNs calculated with LOOCV. All values in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	$Mean_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$Mean_{\text{RMSE, train}}$
ANN	1.0	0.31	1.84	0.64	0.79
ANN$_{\text{Test optimized}}$	0.86	0.26	1.57	0.51	1.30

4.4   Convolutional Neural Network

The following section is divided into three subsections. First, an overview of the scope of the defined CNNs is given, then the hyperparameter optimization is explained. This leads to the impact of the different hyperparameters on model performance. The optimized model is then used to approximate the

generalization performance and finally interpret the model based on Shapley values.

4.4.1 Model Definition and Hyperparameter Optimization

After transforming the images of each tire into 224x224 tensors and matching them with the corresponding acoustic measurements, the ML model can be trained. All images of one tire are combined with each run of this specific tire. The number of tire section images range from 12 to 76. The different amount in these images is due to the masking process (described in more detail in section 3.5) combined with some digitization errors present in the tire circumference images. While the masking process is done for each tire with a pixel step correlating to its specific patch length, the images are afterwards filtered for the appearance of obvious image artifact which leads to the different amount in images per tire. This is done through a manual visual inspection. For each tire, a minimum of two runs, and maximum of six runs is recorded in the speed range of $78.5\,\mathrm{km\,h^{-1}}$ to $81.5\,\mathrm{km\,h^{-1}}$ at PP'. The total data set used in the CNN has 5074 data points for an overall of 125 runs.

Fig. 4.21 presents the schematic structure of the CNN.
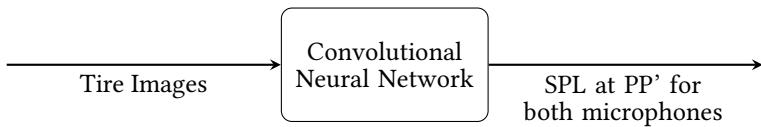
Figure 4.21: Schematic model representation of CNN structure.

Before applying hyperparameter optimization on the CNN, the general expected error is analyzed.

The CNN does not receive any numeric data, not even the vehicle speed, except the SPL values necessary to perform supervised learning. This means

that only one specific vehicle speed can be predicted. Yet it is not possible to only include runs of one exact vehicle speed since this shrinks down the data set to a minimum. Therefore, only runs between $78.5\,\mathrm{km\,h^{-1}}$ to $81.5\,\mathrm{km\,h^{-1}}$ at PP' are included. Assuming an average rolling speed coefficient b of 34.03 dB(A), the expected average difference due to speed can reach $34.03\log_{10}(\frac{81.5}{78.5}) = 0.55\,\mathrm{dB}(A)$. The accuracy in the CNN is therefore not expected to be less than 0.55 dB(A) per microphone, or rather 0.78 dB(A) in RMSE. The speed range around $80\,\mathrm{km\,h^{-1}}$ is chosen since UNECE regulation no. 117, which defines the approval process of tire rolling sound emission, demands measurements in the range of $70\,\mathrm{km\,h^{-1}}$ to $90\,\mathrm{km\,h^{-1}}$ to calculate a reference SPL at $80\,\mathrm{km\,h^{-1}}$ [UNE22].

The hyperparameter optimization includes the following changes in the CNN architecture:

- Activation function in convolutional layer: ReLU, Tanh, Sigmoid
- Convolutional kernel size: 3, 4, 5
- Pooling: Max, Avg, None (kernel size 2)
- Convolution layers: 1, 2, 3, 4,
- Batch size: 4, 16, 64,
- Learning rate: 1e-6, 1e-5, 1e-4,
- Dropout: 0, 0.1, 0.2, 0.4

The number of output channels in the convolution layers is set to twice the number of input channels. Two fully connected layers precede the convolution layers in the CNN. The second fully connected layer breaks down 2^8 input values to two network outputs. After each fully connected layer the values are passed through a ReLU activation function. The dropout layer is situated between both fully connected layers. The hyperparameter optimization receives the loss on training and test data set as criteria for minimization. The models kept are the ones performing best on the test set as well as the TT model (similarly to the process described in section 4.3). Hyperparameter optimization is performed on three independent data splits. As seen in fig. 4.22

the importance of each hyperparameter diverges slightly for different data splits. However the overall trend of importance can be identifies as similar for each split. The as best identified parameter setting varies more gravely. Similar to the observation in section 4.3 in case of the ANN no optimal setting for all three data splits can be identified. The hyperparameter importance is also less stable in case of the CNN than in case of the ANN. For the dropout setting the ten best performing models on the test set identify 0.1, 0.1, and 0.2 through the multi-objective optimization as most promising. However it must be stated that models which perform almost identical have different setting in the dropout layer. This begs the question if the hyperparameter study truly identified true hyperparameter importances or if those are merely impacted by models which did not converge at all.
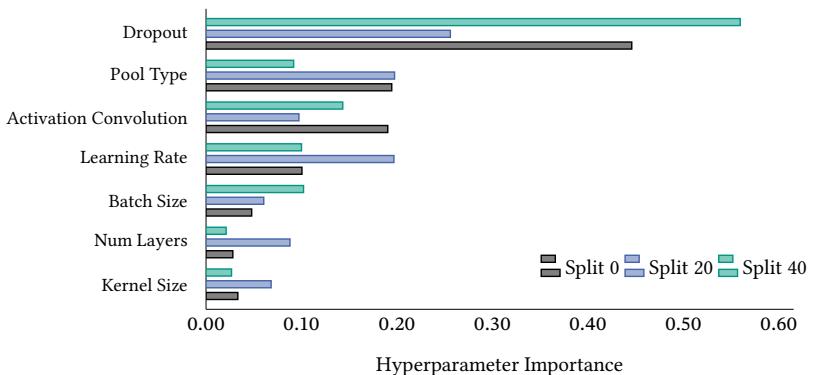
Figure 4.22: Hyperparameter importances for test data set for three data splits on model performance of on the test set for CNNs.

The assumption of the hyperparameter study being largely impacted by outliers is supported by the optimization history presented in 4.17. Identical to the optimization history for the ANN, the different steps in model performance can be identified. It seems that in case of the CNN more models did converge, than in case of the ANN. For the CNN 79/300 reached higher RMSEs than 2 dB(A) while in case of the ANN this was the case for 196/600 models.
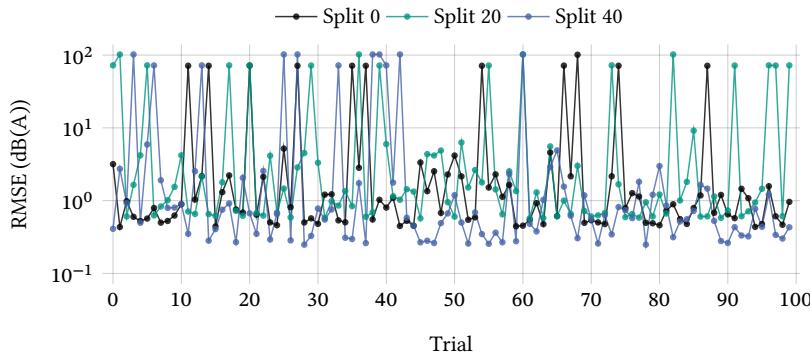
Figure 4.23: Optimization history for three different hyperparameter studies of an CNN. Each study uses a different split for test loss estimation. The figure displays test losses only.
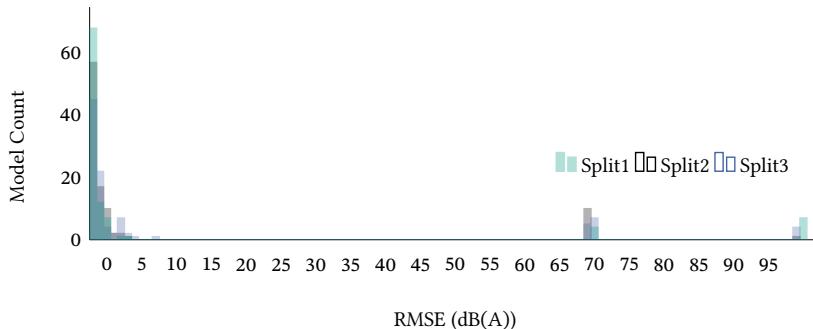
Figure 4.24: Histogram for three different hyperparameter studies of an CNN. Figure displays test losses only.

4.4.2 Performance Evaluation

The model performance is calculated through LOOCV approach with the following settings as trade-off from the hyperparameter study:

- Convolutional activation: Sigmoid

- Kernel size: 3

- Pooling: Max (kernel size 2)

- Convolutional layers: 3

- Batch size: 16

- Learning rate: Adaptive between 1e-5 and 1e-2

- Dropout: 0.1

The learning rate is adapted similarly as in the earlier described ANN after each epoch according to fig. 4.19. Fig. 4.25 presents the losses for each individual split. The corresponding performance is shown in table 4.6. The CNN reaches an overall better mean performance than the ANN, however the deviation between model splits is marginally larger for the CNN approach.
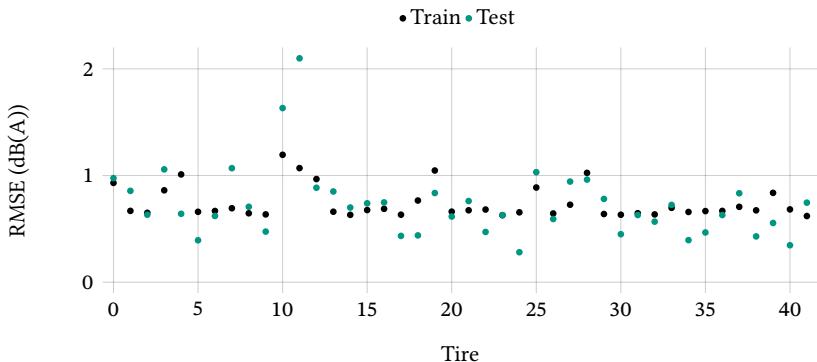
Figure 4.25: RMSE for LOOCV data splits for CNNs defined as TT models.

Table 4.6: Performance metric for CNNs calculated with LOOCV. All values in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	$Mean_{RMSE}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$Mean_{RMSE, train}$
$CNN^{\sim 80\,km\,h^{-1}}$	0.73	0.33	2.10	0.28	0.74
$CNN^{\sim 80\,km\,h^{-1}}_{Test\ optimized}$	0.53	0.17	0.86	0.18	1.67
$ANN^{all\ runs}$	1.0	0.31	1.84	0.64	0.79

As seen in fig. 4.2 the spread in SPLs is around 5 dB(A) for each specific vehicle speed. The restricted data used for the CNN has approximately the same spread. Values for the left microphone range from 70.16 dB(A) to 76.17 dB(A) and 69.79 dB(A) to 76.28 dB(A) for the right microphone. It can be assumed that the performance is, up to a certain amount, defined by the spread in the left out measurements. This is especially the case since each test set only includes two to six different measurements, making it difficult for the CNN to learn to fit their variance especially for both microphone sides, if those tires variance is not represented exactly in the training set. In case of tire eleven as test tire, a RMSE of 2.10 dB(A) is obtained, the true underlying data, as presented in fig. 4.26b, shows that the left out tire includes three measurements whose trend in SPLs for both microphone sides is inconsistent. Furthermore this tire is overall underestimated leading to the worst model split.

The underlying assumption of the images used as input is that the extraction of tread patterns over the circumference of the tire has no influence on the predicted SPL. This means the CNN should predict the same SPL for all individual images of one tire. The prediction for three different data splits (model 3, 11, 24 in fig. 4.25) as well as the corresponding measurement values are shown in fig. 4.26. The difference in the number of predictions is due to the different number of images for each tire as well as the different number of runs. The number of runs can be seen in the steps of the gray and light green curves. In all three models the individual images of one tire do not lead to large differences in the predicted SPLs. Furthermore, the left and right side are almost predicted as identical values, especially in case of fig. 4.26c. However, the important observation is, that each of the models learned to distinguish whether tire fragments are from the same or different tires. This is indicated by the maximum spread between the predictions for the tire fragments of one specific tire and their STDs, which reach 0.78 dB(A) and 0.2 dB(A) for tire 3, 1.32 dB(A) and 0.25 dB(A) for tire 11, and 0.80 dB(A) and 0.17 dB(A) for tire 24. These values are not as low as expected, however compared to the difference in SPL obtained for different tires, it can still be argued that the CNN learned not to distinguish the fragment of one tire gravely.

(a) Data split 3

(b) Data split 11
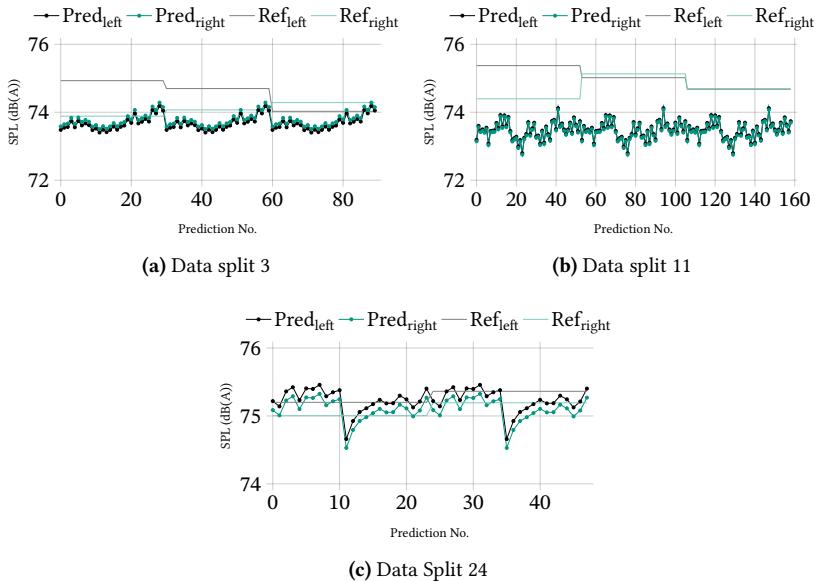
(c) Data Split 24

Figure 4.26: Comparison of model prediction and measurements for three different data splits. Steps in reference values indicate the amount of measurements of the specific tire. Predictions are done for each of the tire digital footprint images of one tire.

The mean RMSE of the CNNs is calculated to 0.73 dB(A) (as seen in table 4.6) and can thus be considered good, especially with the additional information that the mean is pulled upward by only two outliers which reach higher RMSEs than 1.5 dB(A). However, the test error can be interpreted even further. Similar to the training loss due to speed difference which was approximated with an average TRN coefficient at the beginning of this section as 0.55 dB(A), the actual test loss caused by speed differences can be calculated for each tire with the tire specific TRN coefficient. This allows to put the loss caused through inaccurate speed representation in respective to the reached test loss for each data split. Fig. 4.27 shows the corresponding data, yet no clear relation between the losses is visible. The loss due to speed is calculated with the TRN coefficient for each tire averaged for the left and right side and multiplied by $\sqrt{(2)}$ to make its value comparable to the RMSE.
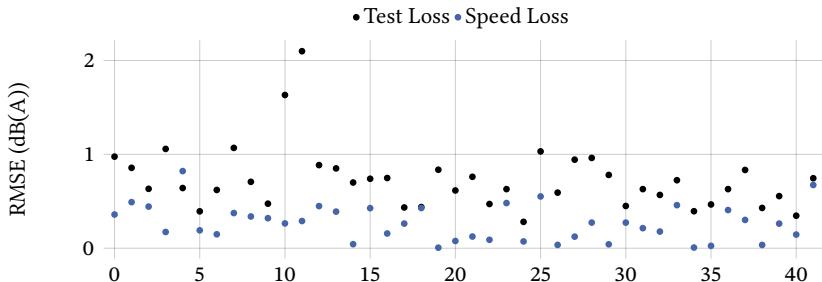
Figure 4.27: Test loss trough inaccurate speed representation compared to actual test loss recorded by each data split.

4.4.3   Model Interpretation

Model interpretation is done with Shapley values on a model which is trained on the whole data set. For this, the DeepExplainer of Shapley Additive Explanations (SHAP) is used since it is computationally designed for DL models and therefore more efficient [Lun17]. As background data, all 5074 images are used. As test data, one image of each tire is picked. Fig. 4.28 shows three different tires as inverted CNN input (4.28a) and their corresponding Shapley values at the left (4.28b) and the right microphone (4.28c). Even though the images do not indicate design rules for less noisy tires, they still allow some insight into the prediction process.

The first observation is the change in the general area of importance. The CNN seems to grasp the shape of the contact patch well and uses this general information as a feature. This is indicated by the overall width and height of colored pixels and the fact that this changes for different tires. Further, it does not seem like the existence of more or wider circumferential grooves strongly affects the predicted SPL. On the other hand, the CNN seems to recognize transversal grooves very well. This is seen in the rows which include blocks. Here, very little rows with constant coloring and therefor impact exist. Rather only the rubber blocks seem to have an impact and the transversal grooves are not categorized as important by the model. This is especially visible in the third image row in the second block row from above. The lowest row shows

all Shapley values for each microphone overlapped, these images are solely included to show that there is in fact a rather large variance in the areas which impact the prediction for each individual tire. All remaining tire footprints and their corresponding Shapley values are available in the appendix D.

(a) Inverted images

(b) Shapley values left microphone

(c) Shapley values right microphone
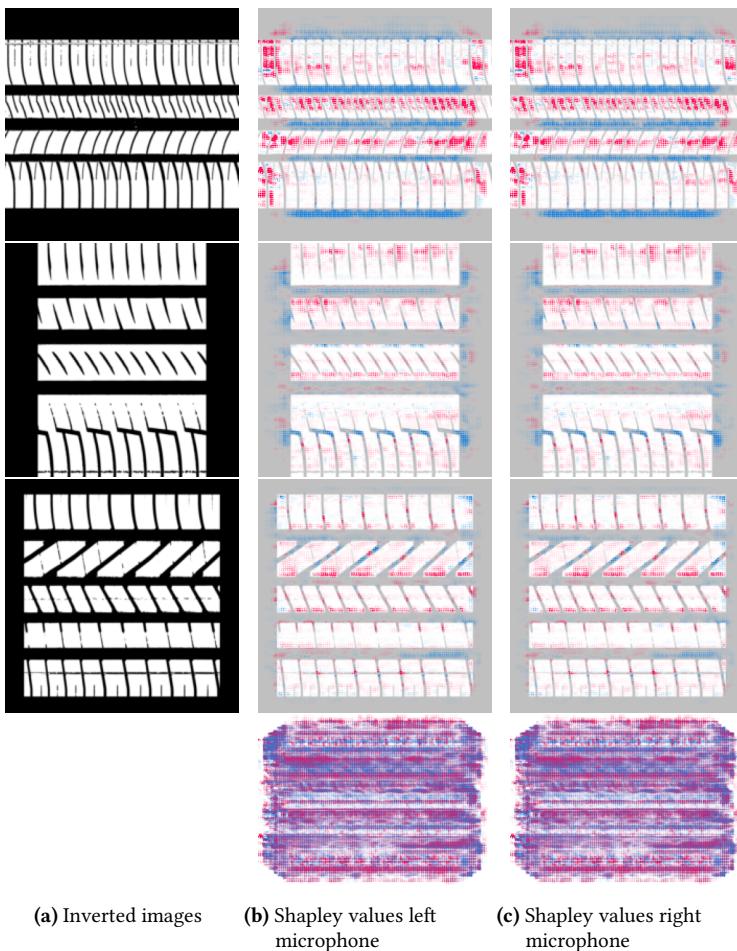
Figure 4.28: Shapley values for images input of CNN structured like reference image, Shapley values left microphone, Shapley values right microphone. Red areas indicate a positive relation between pixel value and the SPL and blue areas indicate a negative relation between pixel value and the SPL. The last row illustrates the overlapped Shapley values of all 42 tires.

4.5 Combination of CNN and ANN

The next logical step in model complexity is a combination of ANN and CNN. The isolated CNN is not able to combine multiple vehicle speeds in one model or take environmental conditions into account. This is especially important considering the possible range in air and road temperature, which is known to have a large impact on the emitted SPL. Also, the temperature range is not as narrowly defined as, e.g., the wind speed so that it could simply be left out. At the same time, the single ANN lacks all relevant information about the tread pattern, from which around 16,000 different versions existed in 2002 [San02].

These observations lead to the idea of combining the CNN with the ANN into one prediction model that receives multi-modal input. In the following, the model which combines ANN and CNN is referred to as Combined Model (CM). Fig. 4.29 shows the principle sketch.
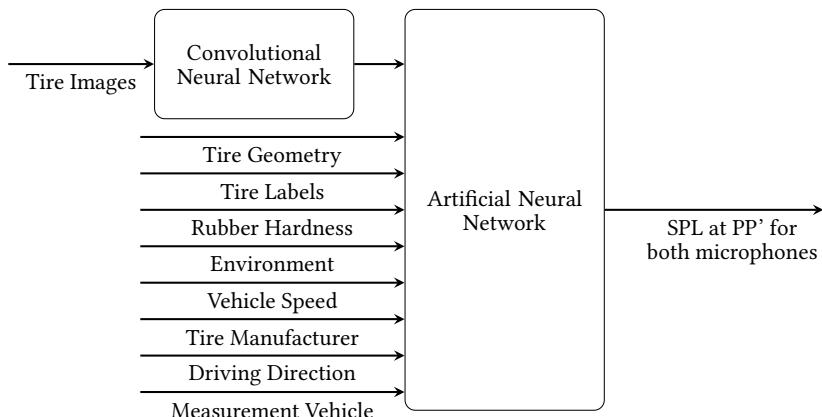
Figure 4.29: Principle sketch of CM. Rubber hardness incorporates two hardness values with correlated surface temperature. Environment incorporates the wind speed and direction, air and road temperature, air pressure, and humidity. Tire geometry incorporates the tire nominal width, the aspect ratio, rim diameter, and profile depth. Tire labels refer to the load and speed index.

Generally, this can be done by either pre-training the model entities by themselves and combining pre-trained models. However, in this case, the ANN receives the output of the CNN as input, which does not allow for supervised learning since no output labels are available for the training of the CNN, nor are all necessary input values available to train the ANN. The output values of the CNN and the embedding layers are combined with all other input values of the ANN instance to one tensor. This is done by simply stacking tensors together to one larger tensor. The ANN receives this tensor as input. The optimizer receives all weights and biases available in the CNN, embedding layers, and the ANN, allowing for the whole architecture to be trained as one entity.

Since the hyperparameter study of the CNN showed that no ideal design existed, the CNN structure is simplified slightly to decrease calculation time. Fig. 4.30 shows the architecture of the used CNN. Hyperparameter optimization is performed for the activation function as well as the layer and neuron count in the ANN structure. This study also includes the learning rate, batch size, and dropout layer probability for 100 different models on three different data splits. As slightly more promising the Elu [Cle15] function as activation function is observed. However, as in the previously shown cases no absolutely clear trends can be derived, which is why it is refrained from discussing the optimization further. The final ANN consists of the input layer, two hidden layers and the output layer.
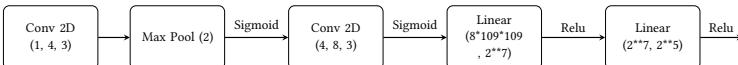
Figure 4.30: Structure of CNN part in combined model approach, nomenclature: Conv 2D(input channels, output channels, kernel size), Max Pool (kernel size), Linear (input nodes, output nodes).

Model performance is analyzed through LOOCV, training and test losses for each split on the TT models are displayed in fig. 4.31. The corresponding metric with comparison to the singular CNN and ANN models is presented in table 4.7. Even though the mean RMSE does not improve gravely and is even worse compared to the CNN approach, the CM has advantages over to

the single models. This is indicated by the smaller STD and especially by the smaller max RMSE which indicates that the CM performs on all data splits better and therefore reaches better generalization than the sole ANNs and CNNs. Furthermore, the mean performance on the training set does not change as gravely as in single models, with 0.74 dB(A) to 1.67 dB(A) for the CNN and from 0.79 dB(A) to 1.30 dB(A) for the ANN, comparing TT best and best test model performance. All data splits are trained for 30 epochs with a fixed learning rate of 1e-4 and no adaptive learning rate since the scheduler showed worse performance in a smaller study. However, model training shows that the process is even after hyperparameter optimization still a stochastic process which means that more epochs or multiple runs of the training process can still improve the model performance. This is visualized in fig. 4.32.
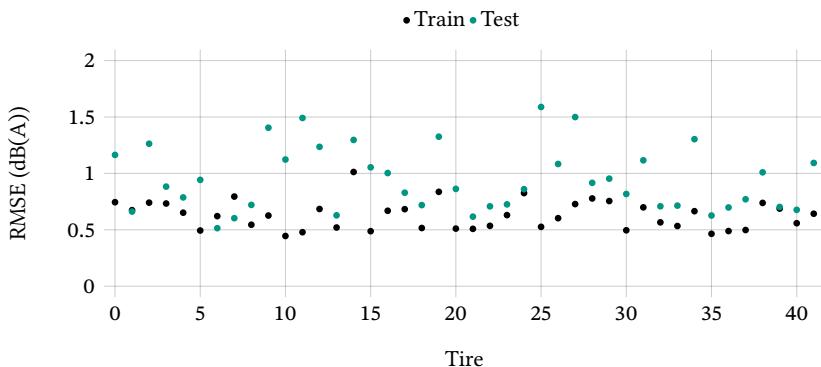
Figure 4.31: RMSE for LOOCV data splits for CM defined as TT models.

Table 4.7: Performance metric for CMs calculated with LOOCV. All values in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	Mean$_{RMSE}$	STD$_{RMSE}$	Max$_{RMSE}$	Min$_{RMSE}$	Mean$_{RMSE, train}$
CM	0.95	0.28	1.59	0.51	0.63
CM$_{Test\ optimized}$	0.82	0.26	1.50	0.49	1.04
ANN	1.0	0.31	1.84	0.64	0.79
CNN$^{\sim 80\,km\,h^{-1}}$	0.73	0.33	2.10	0.28	0.74

(a) Data split 14

(b) Data split 15
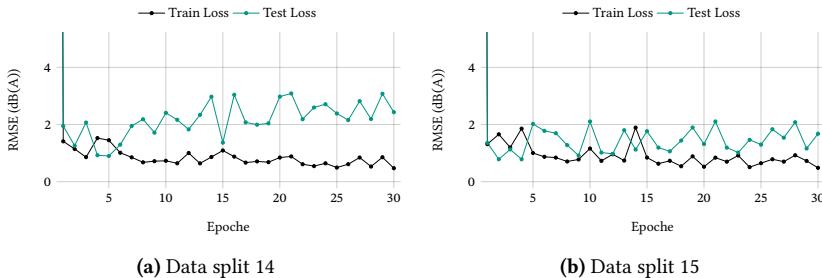
Figure 4.32: Losses per epoch for training and test data set for two individual splits in the LOOCV validation. Epoch zero indicated the losses after initiating the model, before training.

While the loss on the training data set shows convergence for both data splits, the test loss, especially in fig. 4.32a, is much more volatile. This might result from the imbalance in both data sets and the overall small number of data points in the test set. Furthermore, especially looking at the image input, it cannot be verified for each data split if the mathematical operation would represent an extrapolation or an interpolation. The overall fast convergence in the models is not surprising since through the combination of image and numerical data, the CNN receives each image multiple times and the ANN receives each numerical data tensor multiple times. Therefore an epoch actually represents multiple epochs in classic sense of ML problems.

In addition to the TT model performance for the test data set, it is also important to ensure whether the whole data range is covered with reasonable accuracy. Fig. 4.33 displays the predicted and reference values for the test data across two different data splits (for more model representations refer to appendix C). These graphs show that all models seem to capture the impact of the speed accurately, which is indicated by the fact that the predicted SPL increases with increasing reference SPL. Some models tend to slightly over- or underestimate the actual SPL, leading to the TT model losses as recorded earlier.

(a) Data split 14
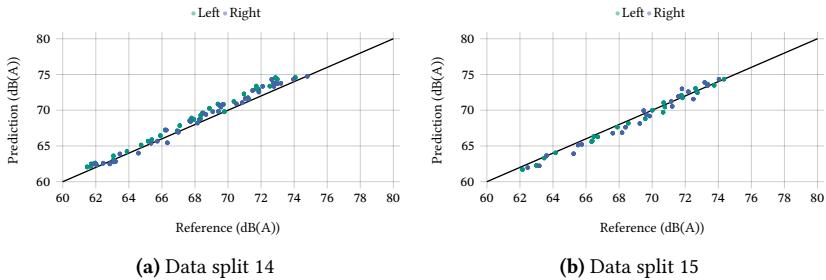
(b) Data split 15

Figure 4.33: Prediction over reference values for two different data splits for test data.

Fig. 4.34 shows measurements which indicate that the shape of the runs of different tires can vary greatly. It can be noted that even repeating runs for the same tire or changing the speed sometimes changes the shape of the measurement curve. The first observation is a result from different radiation characteristics of different tires. How directional the emission characteristic of a tire is, is mainly indicated by the incline and decline at the start and end of the test track, as well as the setback at around 0 m.

(a) Left microphone

(b) Right microphone
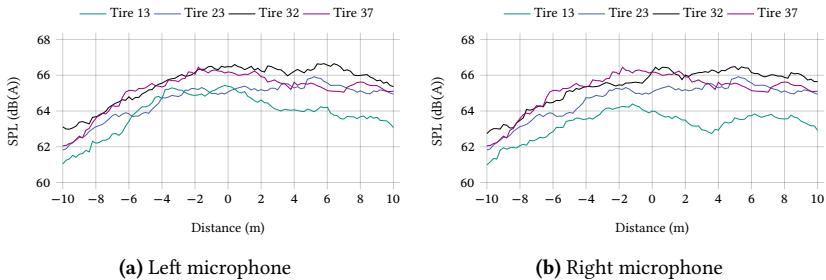
Figure 4.34: Example measurements of idle runs with $50 \, \text{km} \, \text{h}^{-1}$ at PP' showing difference in curve shape for different tires.

This observation makes it necessary to have information about the predicted location on the test track as well. The simplest approach therefore is training an independent model for each point of the test track, which does not need

any adaptation in the networks structure but only requires the input and target values to be updated to the specific location values. The input images remain the same for each location while the numeric data (e.g., vehicle speed or wind direction and speed) change slightly. The difference in the measured SPLs is larger, which is a result of the combination between the distance between the vehicle (sound source) and the microphones as well as the radiation characteristic of the tire. Hereby, the emission direction in driving as well as the opposite direction is dominant over an emission lateral to the rim. This is mainly caused by the horn effect. Even though the time series of the CB measurements are fitted to the track distance in steps of 0.2 m, model training is only carried out for every 0.4 m step between -10 m and 10 m. This leads to 51 models to cover the entire track, with three splits of the LOOCV approach 153 models are trained in total.

Fig. 4.35 shows the model performance on the test set for these three data splits over the whole distance. The figure indicates that no point of the track seems to be generally worse in the prediction for all three investigated tires. However, the range in minimum and maximum predictive accuracy over the test track slightly changes for the three data sets. This is indicated trough maximum and minimum performance in the performance metric in table 4.8. This still does not indicate that these models generalize worse than the CM for only position PP'. The slightly higher RMSE can also be caused by higher variance in the underlying measurements, especially in higher spreads between the left and right microphones.
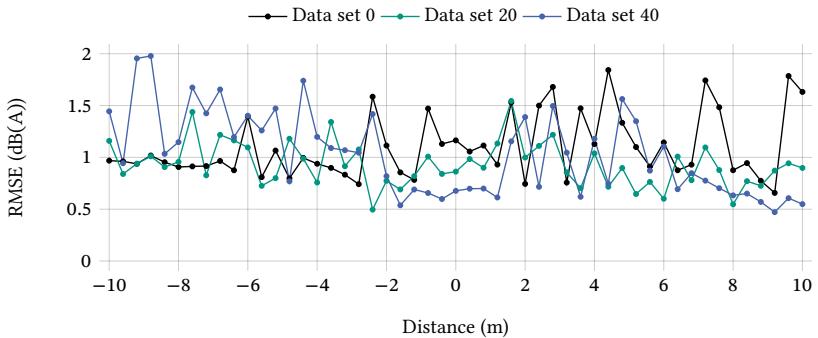
Figure 4.35: RMSE for TT model performance on test data set for three different data splits of CM over the whole test track.

Table 4.8: Performance metric for CMs for three data splits over the whole test track calculated with LOOCV. All values in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	$Mean_{RMSE}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}
Data set 0	1.10	0.34	1.84	0.66
Data set 20	0.93	0.23	1.54	0.50
Data set 40	1.03	0.44	1.98	0.47

Fig. 4.36 shows that compared to the measurement used as ground truth, the concatenated predictions over the distance look way less stable. The higher volatility between two predictive steps is amplified by the larger steps between two data points for the prediction compared to the measurements. Also, since models are trained independently, the model cannot learn any relation for the distance, which makes it more unlikely to get a smooth curve as suggested by the measurements.

(a) Tire 0

(b) Tire 20

(c) Tire 40
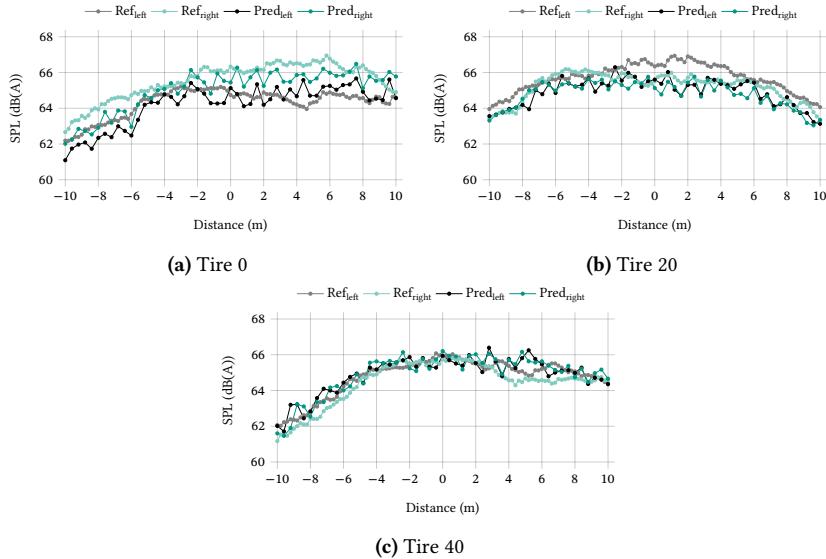
Figure 4.36: Example measurements of idle runs with $50\,\mathrm{km\,h^{-1}}$ at PP' showing difference in comparison to predicted values of independent models for each distance. Predictive step is $0.4\,\mathrm{m}$ and measurement step is $0.2\,\mathrm{m}$.

It can also be noted that the prediction over the distance also inherits some error. The performed measurements are in classical sense time series. Through the combination of the known start of the measurement at position AA' and the speed measurement, the time steps can be interpolated on the test track. Since this does not guarantee a measurement point each $0.2\,\mathrm{m}$, the used measurement software interpolates the SPL to these points. However, it is not probable that the variance included through this process causes a noticeable impact on the measurement. If this led to a noticeable error, the expected error would grow with increasing vehicle speed. This results from the decrease in measurement time to pass trough AA' and BB'. Through the defined sampling frequency, the grid in the SPL over distance curve gets wider and thus the impact of interpolation two discrete points is larger.

4.6 Combination of CNN and ANN over Distance

As section 4.5 showed, it is possible to predict the SPL over the whole test track with independent models for each point. However, through this the relationship between the data points is lost and it is less likely that the predictions will be smooth. To avoid this information loss, in the following a CM is defined which enables predictions over the full test track. Slightly modified versions of this model are then used for model interpretation.

4.6.1 Model Definition and Performance Evaluation

Basically, the same model as the CM which only predicted one point on the test track, can be trained for the purpose of predicting the whole CB measurement. As illustrated in fig. 4.37 the only change necessary considering the model architecture is to enlarge the ANN input by one dimension, the distance. This makes changes in the data structure necessary. The entire data set consists of 1443 runs, each with 101 points on the test track, resulting in a total of 145743 data points. Merging every tire image with the connected data points, as done in previously presented models, is not reasonable anymore. Therefore, each data point of a run gets one cropped tire image assigned. During this process, it is ensured that each of the cropped images of the specific tire is represented equally often in the data set. This process does not enlarge the whole data set. Equally to the CM without distance information, the input data are scaled with a z-score normalization as described in eq. (3.2). All numerical input values except the distance are scaled. Since the distance is already a uniformly distributed variable through the fixed step of 0.2 m between $-10\,\text{m}$ to $10\,\text{m}$, it is not expected that the transformation of the distance variable will increase the model performance.
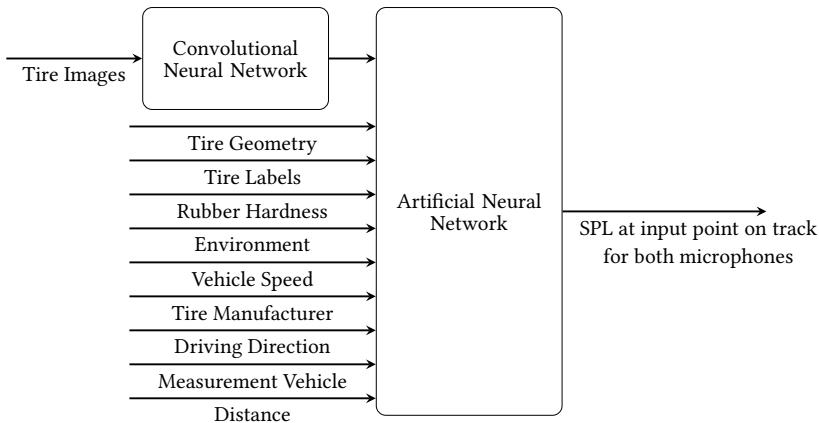
Figure 4.37: Principle sketch of CM$^{\text{Distance}}$. Rubber hardness incorporates two hardness values with correlated surface temperature. Environment incorporates the wind speed and direction, air and road temperature, air pressure, and humidity. Tire geometry incorporates the tire nominal width, the aspect ratio, rim diameter, and profile depth. Tire labels refer to the load and speed index.

Fig. 4.38 shows the losses received for each model through LOOCV. The test performance is more stable than in all previously built models, since the lowest STD for RMSE of all models considered so far is reached. The train performance is also visibly more stable.
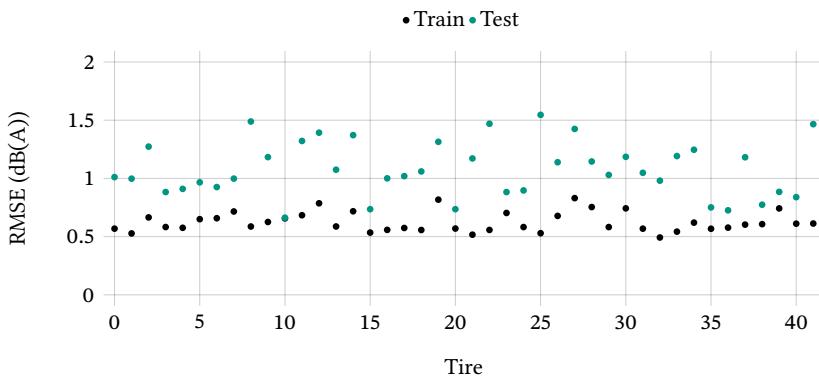
Figure 4.38: RMSE for LOOCV data splits CM distance models defined as TT models.

This is further supported by table 4.9 which shows the performance of the distance model in comparison to the previous approaches. While the mean performance on the RMSE of the TT models slightly deteriorates compared to the single point CM, the STD of the RMSE as well as its maximum value decreases. However a direct comparison of the mean RMSE of these models does not seem reasonable. This arises from two observations. Firstly, even though the spread between the microphones on both sides is not gravely impacted by the distance on the test track (as seen in fig. 4.3b), this does not indicate if the spread between quasi-identical runs is larger at some points of the test track. Secondly, including every position of the test track increases the number of data points and therefore the likelihood of predicting points with low accuracy, which affects the performance of the model.

The first argument is validated by the observation in table 4.8 that the purely stacking of models for each point on the test track in three cases leads to almost no change in the mean RMSE performance. However, in two of three cases, the maximum prediction error gets worse.

Table 4.9: Performance metric for CMs for full test track calculated with LOOCV. All values in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	$\text{Mean}_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$\text{Mean}_{\text{RMSE, train}}$
$\text{CM}^{\text{Distance}}$	1.08	0.23	1.55	0.66	0.62
$\text{CM}^{\text{Distance}}_{\text{Test optimized}}$	0.98	0.21	1.55	0.60	1.00
$\text{CM}^{\text{PP'}}$	0.95	0.28	1.59	0.51	0.63
ANN	1.0	0.31	1.84	0.64	0.79
$\text{CNN}^{\sim 80\,\text{km h}^{-1}}$	0.73	0.33	2.10	0.28	0.74

Fig. 4.39c displays the identical three runs as earlier predicted by the stacked models for each individual distance (fig. 4.36). The full distance model produces much smoother SPL curves for each of the three runs compared to the previous prediction and also in comparison to the actual measurements. It can also be stated that even though the location of the maximum SPL for tire 20 is not fitted correctly, the shapes in measurement and simulated curves

seem to align rather well. Tire 0 and tire 40 raise rather steeply until point PP' in the measurement as well as in the simulation and are rather constant after that point. In tire 20 the measurement shows a steep decline in SPL after its maximum, this is also indicated by the simulation. This implies that the model learns some underlying radiation characteristics for each tire from the data set. To analyze the radiation pattern of the tires based on the angle between tires and microphones, the vehicle length must also be considered. This results from the measurement process used in measurements conducted for this dissertation, which triggers the measurement the moment the front of the vehicle crosses position AA' and stops the measurement, the moment the vehicle front passes position BB'. Taking this into account, the steep incline in SPL for tire 0 up to position PP' combined with its plateau afterwards indicate a stronger signal amplification through the horn effect at the trailing edge than on the leading edge. Tire 20 seems to have the exactly opposing radiation pattern. While the SPL at entrance and exit of the measurement track is rather similar, which would indicate a radiation pattern which is similar for the trailing and leading edge, however taking the measurement vehicle length in consideration, the tire indicates a stronger amplification on the leading edge than on the trailing edge. Tire 40 shows the in between, the measurements drop slightly more towards the end of the measurement track compared to tire 20. Tire 40 therefore indicates a more or less identically strong signal amplification for the leading and trailing edge.

(a) Tire 0

(b) Tire 20

(c) Tire 40
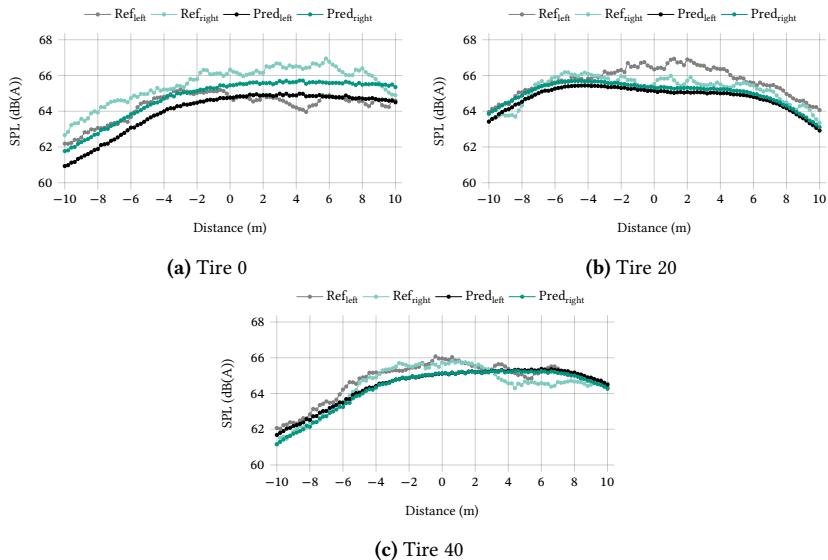
Figure 4.39: Example measurements of idle runs with $50 \, \mathrm{km \, h^{-1}}$ at PP' showing difference in comparison to predicted values of distance dependent models. Predictive step is $0.2 \, \mathrm{m}$ and measurement step is $0.2 \, \mathrm{m}$. Predictions are performed for test data.

Using every data point for model training leads to a RMSE of $0.50 \, \mathrm{dB(A)}$. Even though this model cannot be used to analyze the model's generalization capability, it is still useful to analyze the model's capability to represent the data in general. This allows to estimate if the model complexity is high enough for the specific problem.

Consequently, fig. 4.40 shows six different measurements of six different tires with their belonging prediction of $CM^{Distance}$ trained on the whole data. The figure shows once more the variability in the overall shape of the measured SPL curves for different tires which makes it a more complex problem for the model to learn. Also, the predicted SPLs are much less volatile compared to the real measurements, making it less probable for the model to reach RMSE values near zero. This, however, indicates that even after using the whole data set in training, the data points are not simply remembered by the network, yet it cannot be ruled out that the model learns specific shapes for each tire. The

network also learns to predict different values for both microphone sides. In the shown examples, the prediction for the right microphone is always slightly higher, which is also indicated by the measurement.

(a) Tire 9

(b) Tire 10

(c) Tire 15

(d) Tire 28

(e) Tire 36

(f) Tire 41
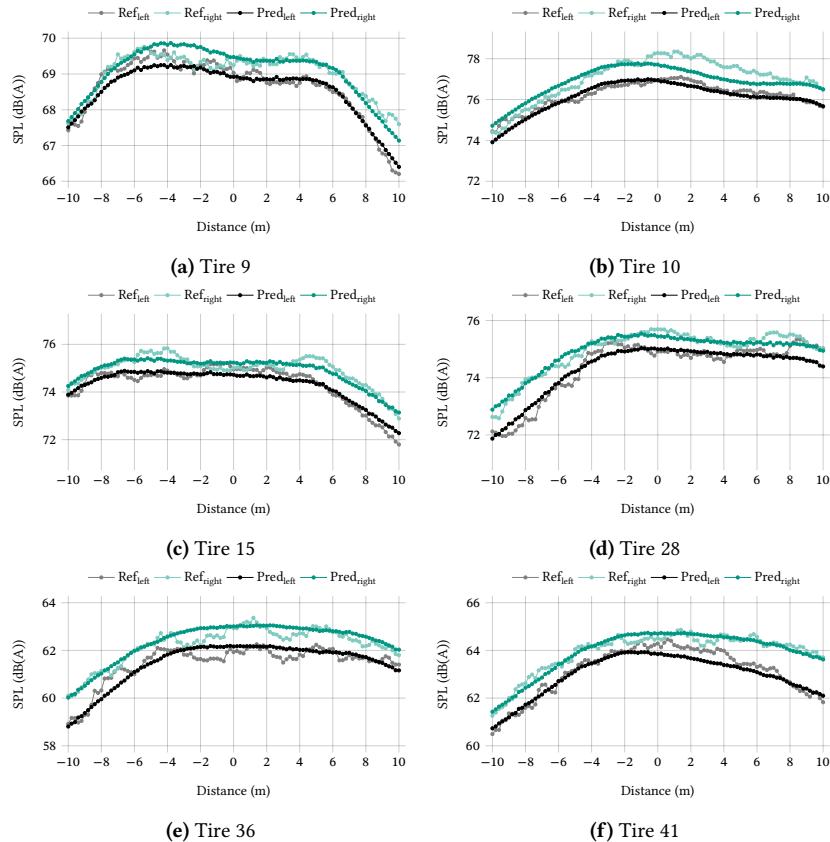
Figure 4.40: Example idle runs with corresponding prediction of $CM^{Distance}$ with full data set for training. Predictions are performed for training data,

4.6.2 Model Interpretation

The model trained on the whole data reproduced the measurements accurately. This is a necessary condition to derive information through model interpretation. However, it is reasonable to change the model input before applying XAI methods on it. This is due to the rather high correlation between some of the features of the models as seen in fig. 4.6. A more profound insight into the multicollinearity of the features can be seen through the VIF. For four different models, fig. 4.41 shows the correlating VIF values as obtained for the actual data set used in the CM. $CM^{Distance}$ refers to the CM as described in section 4.6. All other models do not include the humidity and the road temperature. In case of the temperatures, the VIF can mainly be reduced by only using either the road or the air temperature in the model. Since the air temperature is generally more stable due to less local effects like the sun radiation, the air temperature is kept as input. Model $CM^{Distance}_{tire\ aspect\ ratio}$ also does not receive the tire aspect ratio. $CM^{Distance}_{rim\ diameter}$ does not receive the rim diameter. $CM^{Distance}_{outer\ diameter}$ does not receive the tire aspect ratio, the rim diameter, and the tire load index. Instead of these $CM^{Distance}_{outer\ diameter}$ receives the outer diameter of the tire which is a combination of the tire nominal width, tire aspect ratio, and rim diameter. The VIFs drop considerably for all model variations. Only $CM^{Distance}_{rim\ diameter}$ has slightly higher values for the tire nominal width and the tire aspect ratio.
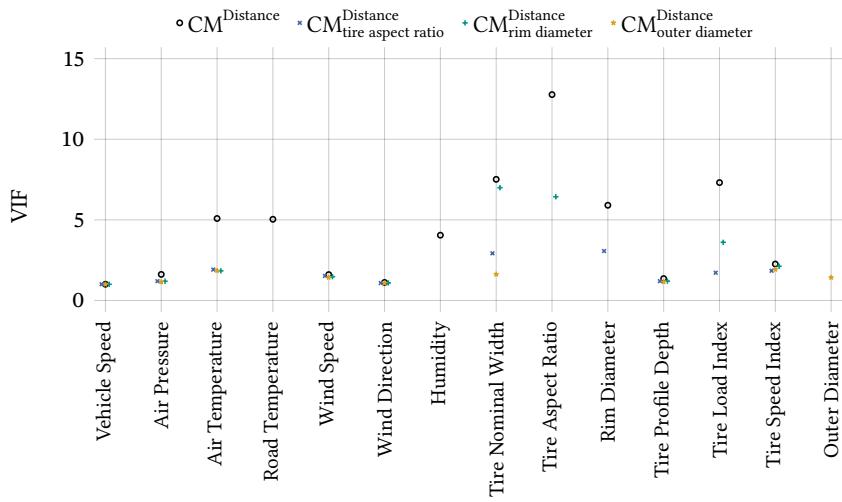
Figure 4.41: VIF values of model variations calculated for all features passed directly to the model structure without being preprocessed in an embedding layer. To enhance the visibility of multiple data points that have the same values, the markers are slightly shifted along the x-axis.

Table 4.10 shows the performance metric of each of the models in comparison to the model which included all features.

Table 4.10: Performance metric for CMs for full test track calculated with LOOCV for multiple data input manipulations. All values in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 2 elements each.

Model	$\text{RMSE}_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$\text{Mean}_{\text{RMSE, train}}$
$\text{CM}^{\text{Distance}}_{\text{tire aspect ratio}}$	1.14	0.22	1.57	0.65	0.62
$\text{CM}^{\text{Distance}}_{\text{rim diameter}}$	1.17	0.24	1.64	0.79	0.60
$\text{CM}^{\text{Distance}}_{\text{outer diameter}}$	1.20	0.24	1.69	0.74	0.63
$\text{CM}^{\text{Distance}}$	1.08	0.23	1.55	0.66	0.62

Since these models show high accuracy and no critical multicollinearity, they are suitable for model interpretation. As an alternative to the use of SHAP

package to estimate shapley values, Captum (presented in section 2.5.1) provides a convenient method for estimating feature importance of multi modal models based on *Integrated Gradients*. For the use of this method, an explicit basevalue must be defined. This basevalue should not inherit any information. In case of the tire images this is somehow difficult to define. Yet the usage of a completely black image as background image provides advantages. The black image thereby represents a slick tire. Through the definition of the basevalue, the algorithm gets insensitive to these values. In this specific investigation that means, black pixels do not impact the estimated feature attribution. This is interesting since it actively allows to analyze the impact of any grooves on the models prediction.

To investigate the impact of the image input of the tire on the prediction, the following approach is applied to the data and the model. First, the embedding layers are deactivated and the corresponding values of variables that use embedding layers are precalculated. Second, the basevalues are defined as minimum values in each numerical category and the tire is defined as slick.

For this, the cropped tire image is set to black for its entire image fragment before it is pasted on the white background, as described in section 3.5. As input values for calculation of the IG, 1000 random data points of each tire are used. These consist of full sets of data including numerical and image data as specified through the necessary input for the ML model. While the numerical data are kept as randomly selected, the image data are manipulated. This is done in a way that each of the 1000 data points of one tire has the same tire fragment. The used fragment is chosen randomly. This is necessary to avert a blurring effect in the influence of vertical grooves when building the mean impact over all images of one tire. This occurs because of the horizontal shift of said grooves in these images. This process allows for an overall evaluation of the tire tread pattern on the emitted noise.

Furthermore, a second set of basevalues is defined with all tires reconstructed as semi slick. This is achieved by first dilating over the cropped images to eliminate small outliers in circumferential grooves and afterwards eroding over the image with a kernel of almost no height but great width before pasting it on the white background. This allows to leave the width of the circumferential

blocks nearly untouched but to close all included grooves of any orientation not equal to 0°.

Fig. 4.42 shows a full set of images for the investigation of feature importance.

(a) Original tire cropped image   **(b)** Tire slick representation   **(c)** Tire semi slick representation
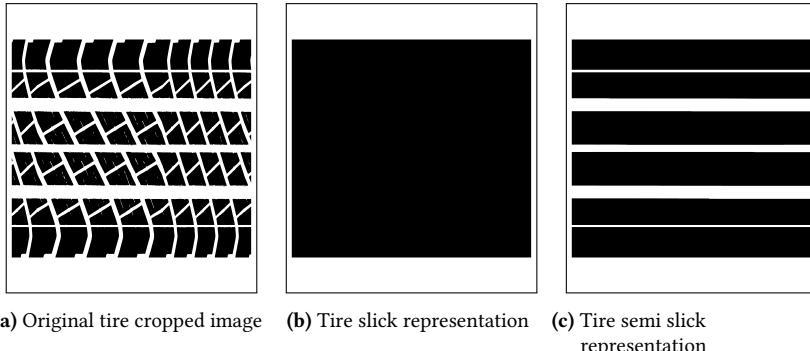
Figure 4.42: Three different tire images after cropping and rescaling for standardizes CNN input with each pixel size 3947 x 4559. Frames only added for displaying purpose

The IGs do not show clear trends for the images. This was already seen in the model interpretation of the CNN through SHAP. Fig. 4.43, fig. 4.44, and fig. 4.45 show two tires as input compared to the slick and semi slick representation for both output microphones for each of the adapted models. The scale is modified to be symmetric and chosen as the absolute maximum pixel representation divided by ten. This is necessary for visual representation. It can be pointed out that trends like circumferential grooves leading to lower SPL are not supported by the images. Neither do higher angles in transversal grooves show any clear trend. Also, no noticeable changes between the different models can be observed. The displayed values in the figures represent the average output for the 1000 random samples which are explained for each tire. The full set of images is displayed in appendix E.

(a) Tire 23 semi slick left

(b) Tire 23 semi slick right

(c) Tire 23 slick left

(d) Tire 23 slick right

(e) Tire 38 semi slick left

(f) Tire 38 semi slick right

(g) Tire 38 slick left

(h) Tire 38 slick right
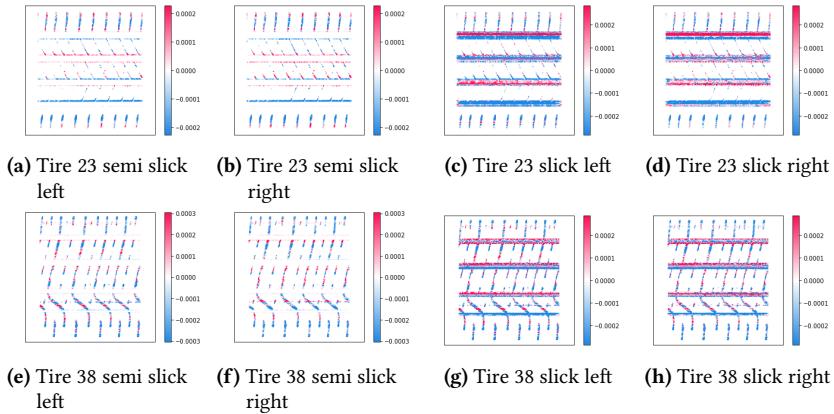
Figure 4.43: Feature importance interpretation through IG for baselines as slick and semi slicks. Numerical values are chosen as minimum baseline. Left and right in sub captions refer to the output feature, i.e., the microphone. Color scales are defined by a tenth of the absolute maximum value occurring in a pixel and defined symmetrically. Blue pixels indicate negative correlation to the SPL, red pixels indicate positive correlation with the SPL. As Model $CM_{\text{tire aspect ratio}}^{\text{Distance}}$ is used.

(a) Tire 23 semi slick left

(b) Tire 23 semi slick right

(c) Tire 23 slick left

(d) Tire 23 slick right

(e) Tire 38 semi slick left

(f) Tire 38 semi slick right

(g) Tire 38 slick left

(h) Tire 38 slick right
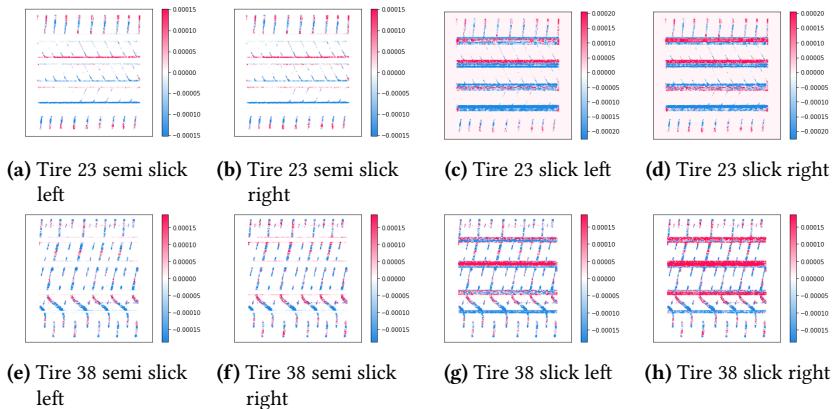
Figure 4.44: Feature importance interpretation through IG for baselines as slick and semi slicks. Numerical values are chosen as minimum baseline. Left and right in sub captions refer to the output feature, i.e., the microphone. Color scales are defined by a tenth of the absolute maximum value occurring in a pixel and defined symmetrically. Blue pixels indicate negative correlation to the SPL, red pixels indicate positive correlation with the SPL. As Model $CM_{\text{rim diameter}}^{\text{Distance}}$ is used.

(a) Tire 23 semi slick left

(b) Tire 23 semi slick right

(c) Tire 23 slick left

(d) Tire 23 slick right

(e) Tire 38 semi slick left

(f) Tire 38 semi slick right

(g) Tire 38 slick left

(h) Tire 38 slick right
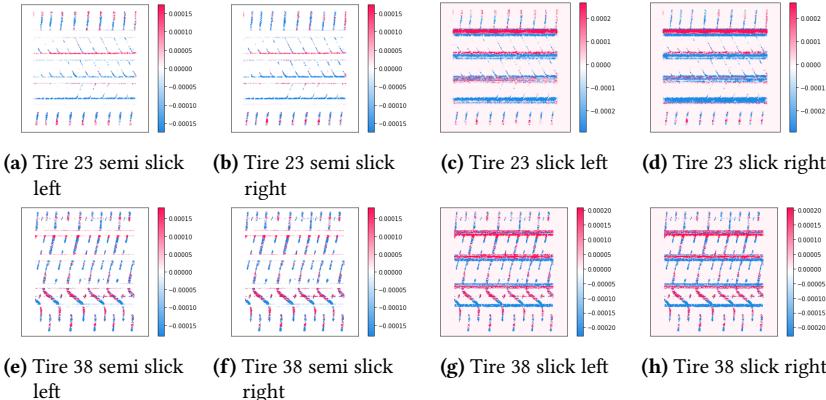
Figure 4.45: Feature importance interpretation through IG for baselines as slick and semi slicks. Numerical values are chosen as minimum baseline. Left and right in sub captions refer to the output feature, i.e., the microphone. Color scales are defined by a tenth of the absolute maximum value occurring in a pixel and defined symmetrically. Blue pixels indicate negative correlation to the SPL, red pixels indicate positive correlation with the SPL. As Model $CM_{\text{outer diameter}}^{\text{Distance}}$ is used.

Fig. 4.46 shows the interpretation of the numerical input. The vehicle speed is for all modified models the most important feature. The air temperature is determined as the feature with the second largest negative correlation to the SPL. In case of the feature which shows the strongest negative correlation with the SPL, the models are also identical. Models $CM_{\text{rim diameter}}^{\text{Distance}}$ and $CM_{\text{tire aspect ratio}}^{\text{Distance}}$ which include the tire load index determine the tire load index as said feature, while model $CM_{\text{outer diameter}}^{\text{Distance}}$ determines the outer diameter. However, this is not contradictory since $CM_{\text{outer diameter}}^{\text{Distance}}$ does not include the tire load index and the tire load index shows a strong correlation to the outer diameter (see fig. A.1). This allows to argue, that it might be able to simply swap the tire load index and the outer diameter, as indicated by the IG. The tire nominal width, distance, and the tire profile depth always show a positive correlation with the SPL, however their absolute impact does not remain constant. It can also be observed that the impact of the tire aspect ratio in model $CM_{\text{rim diameter}}^{\text{Distance}}$ is almost identical to the impact of the rim diameter in model $CM_{\text{tire aspect ratio}}^{\text{Distance}}$.

(a) $CM_{\text{tire aspect ratio}}^{\text{Distance}}$

(b) $CM_{\text{rim diameter}}^{\text{Distance}}$

(c) $CM_{\text{outer diameter}}^{\text{Distance}}$
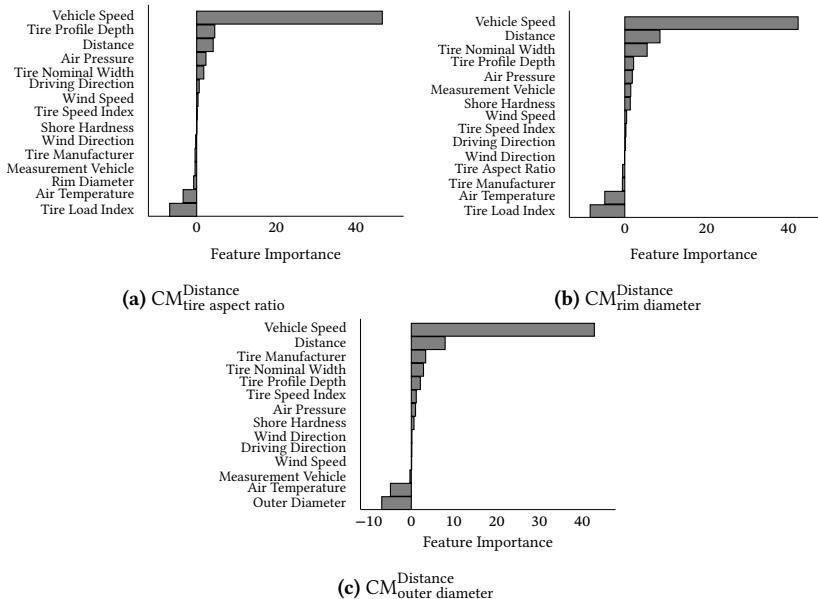
Figure 4.46: Comparison of feature importance obtained through IG for three models with feature adaptions.

Model explainability or feature importance is also possible through directly looking into changes of the predictive value caused by iteratively changing one single feature. However, this neglects interdependencies between all features. This problem can be overcome by randomly sampling a fixed value of data points first and second iteratively changing only the feature under investigation in the sampled data. The predicted value can thus be averaged to analyze overall trends in the underlying data. This is repeated for each feature independently. By choosing a reasonably large number of samples, it is not to assume that only local trends are reproduced. The process relies on accurate inference of the trained model. This is especially the case for models which are trained on the whole data set, which is why these models are used to identify overall data trends. This procedure is especially useful since it basically allows to break down multidimensional problems to a single input feature and its related output variable but yet takes all other features into

account, thus making it possible to visualize trends which are not displayed by simply displaying feature values over labels of the data set. This process is performed for each of the values tire nominal width, tire aspect ratio, tire speed index, tire load index, rim diameter, vehicle speed, driving direction, distance, tire manufacturer, measurement vehicle, air temperature, and air pressure independently. For the one-hot encoded values measurement vehicle, tire manufacturer, and driving direction the procedure is slightly adapted. The encoding allows for an easy adaptation of the values for all the samples under investigation. This means, in case of the direction, all chosen samples are first manipulated to enter the test track from one side and afterwards from the other side. The thus obtained values are subtracted from another to receive a delta for each of the microphones caused by the driving direction. Similarly for the measurement vehicle as well as the tire manufacturer, the input samples are also manipulated in such a way, that each sample is considered as each of the measurement vehicles and tire manufacturer once.

The analysis of the impact of the vehicle speed on the SPL in fig. 4.47a shows the behavior as expected for model $CM_{\text{rim diameter}}^{\text{Distance}}$. The data are well represented by a logarithmic curve. This again supports the speed SPL relationship frequently described in the literature. The distance variation in fig. 4.47b reveals how directional the average tire's sound emission actually is. More interestingly, it also indicates that the tires under investigations on average have an emission characteristic that is more prominent on the front side of the tire than the rear side. This is shown by the fact that the SPL reaches its maximum value before it passes the microphones and continues to fall monotonously from then on.

(a) Vehicle Speed

(b) Position on Test Track
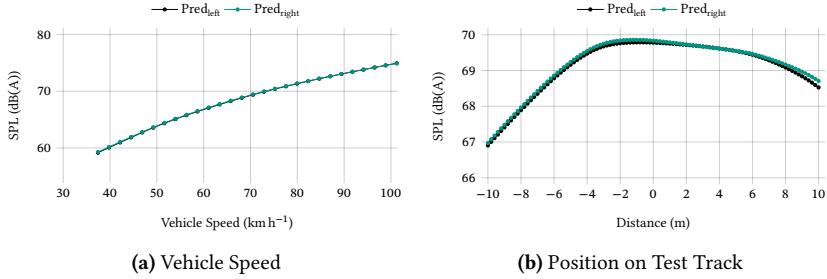
Figure 4.47: Feature importance of $\text{CM}_{\text{rim diameter}}^{\text{Distance}}$.

In case of the tire parameters shown in fig. 4.48, similarly to the investigation with IG, the iterative process reveals the tire nominal width, tire profile depth, and the tire load index as most important for $\text{CM}_{\text{rim diameter}}^{\text{Distance}}$. All indicate linear behavior. In case of the tire nominal width, this was expected from previous literature, however some literature considered narrower tires and came to the conclusion that for wider ones the impact of tire nominal width is negligible [San02, Yan13]. This is not supported by the model interpretation determined here, in which the linear effect remains intact for the investigated range of tire nominal widths from 205 mm to 285 mm. The impact of the tire nominal width and the tire profile depth for the other models are comparable to the ones shown in fig. 4.48. The impact of the tire load index is also comparable to that of the outer diameter. In case of the tire aspect ratio and the rim diameter, the trends are not comparable. However as seen in fig. 4.48b the tire aspect ratio also does not show a clear trend in the one model in which it is included. The corresponding diagrams for the other model modifications are displayed in appendix F.

(a) Tire nominal width

(b) Tire aspect ratio

(c) Tire speed index

(d) Tire load index

(e) Tire profile depth
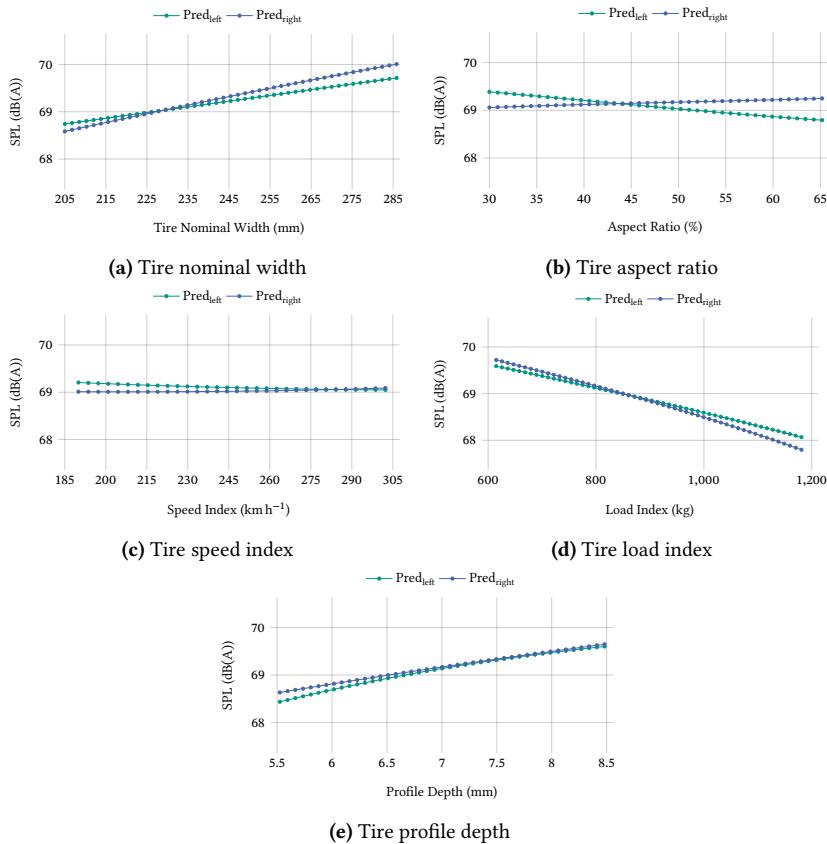
Figure 4.48: Feature importance of $\mathrm{CM}_{\mathrm{rim\ diameter}}^{\mathrm{Distance}}$.

In case of the impact of the air temperature and the air pressure, the results are also consistent with those received with IG. The air temperature shows a strong negative correlation with the SPL, which is also what the general sentiment from literature shows. In case of the air pressure the relation is positive. However, an effect of the environmental air pressure on the TRN is not recorded elsewhere in literature, known to the author. The impact of the air temperature is obtained as $-1.63\,\mathrm{dB(A)}$ for the maximum recorded air temperature difference of $18.06\,°\mathrm{C}$ when both microphones are averaged. As

suggested earlier, with the MB temperature correction applied to air or road temperature the expected difference reaches $-0.95\,\text{dB(A)}$ and $-1.27\,\text{dB(A)}$ so that the model defined the temperature impact as more important than the predefined temperature correction.

(a) Air temperature

(b) Environmental air pressure
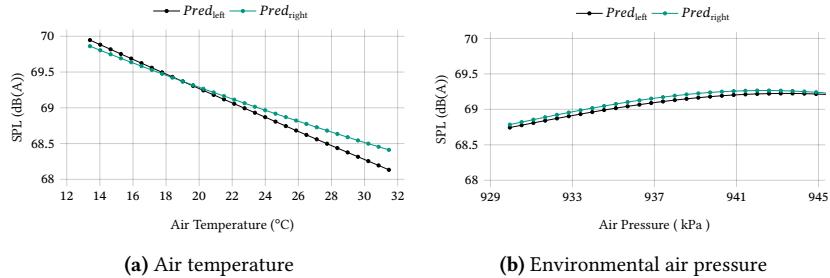
Figure 4.49: Feature importance of $\text{CM}_{\text{rim diameter}}^{\text{Distance}}$.

Fig. 4.50 shows the results for the one-hot encoded features. Since the driving direction only has two possible values, from left to right and from right to left, fig. 4.50a displays the difference in output values. The mean values for the left and right sides are marked in red at $-0.27\,\text{dB(A)}$ and $0.16\,\text{dB(A)}$. Due to the symmetry of the measurement process, values close to zero would be expected. However, reasons as e.g. slight cross or longitudinal slope of the test track as well as local asphalt effects might lead to these effects. The slight divergence might also be caused by unbalanced data since not exactly the same amount for drives from both sides exist in the data. In case of the tire manufacturer in fig. 4.50b the predictions for the tires are in the range of $-0.15\,\text{dB(A)}$ to $0.29\,\text{dB(A)}$ from the mean. For the measurement vehicle this spread is in the range from $-0.40\,\text{dB(A)}$ to $0.69\,\text{dB(A)}$. These ranges are slightly different for the other investigated models as shown in appendix F. Generally, the analyzed impact of the tire manufacturer as well as the measurement vehicle should be viewed with caution. This is the result of the unbalanced data when it comes to these variables since not the same number of tires of each of the 14 tire manufacturers is used. The same applies for the vehicles. The majority of tires is measured on vehicle 1 and 2.

(a) Driving direction

(b) Tire manufacturer

(c) Measurement vehicle
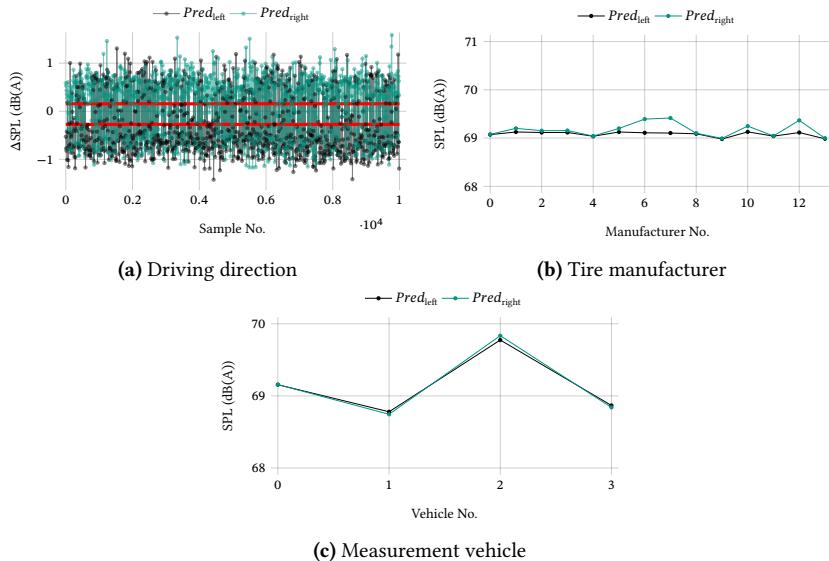
Figure 4.50: Feature importance of one-hot encoded features of $CM_{\text{rim diameter}}^{\text{Distance}}$.

4.6.3 Model Averaging Outputs and Tread Pattern Variation

As already shown in fig. 4.3b, the deviation between both microphones used during the measurements can be rather high even though it seems to be Gaussian distributed. This is also a result of the different tire geometries and structures which are differently designed so that each of the tires might show a different behavior in the spread of emitted sound to both sides. This, combined with the fact that the certification regulations take the maximum SPL of each run, no matter the microphone which recorded it, justifies using the average SPL of both microphones as output of the ML model.

This does not require any changes in the model itself, so that the model still refers to the principle sketch in fig. 4.37 with only the output adapted from vector of the shape (y_1, y_2) to $\frac{y_1+y_2}{2}$. Table 4.11 shows the specific performance metric.

Table 4.11: Performance metric for CMs for full test track with average value of both microphones calculated with LOOCV. All values in dB(A). RMSE values of third row are calculated for an output vector of dimension two.

Model	Mean$_{\text{RMSE}}$	STD$_{\text{RMSE}}$	Max$_{\text{RMSE}}$	Min$_{\text{RMSE}}$	Mean$_{\text{RMSE, train}}$
CM$_{\text{Avg}}^{\text{Distance}}$	0.64	0.16	0.93	0.34	0.39
CM$_{\text{Avg ,Test optimized}}^{\text{Distance}}$	0.60	0.16	0.92	0.28	0.50
CM$^{\text{Distance}}$	1.08	0.23	1.55	0.66	0.62

Due to the different RMSE criterion, the values of the model predicting both microphones independently (CM$^{\text{Distance}}$) need to be divided by $\sqrt{(2)}$ for comparison, which leads to a mean RMSE of 0.76 and a max value of 1.09. These values are still slightly higher than the ones recorded for the averaged model.

Even though a mean RMSE of 0.64 dB(A) is rather low, using the tire fragments as rectangles is a rather strong simplification of the underlying physics. This is why the Fuji pressure foils, as described in section 3.4.2, are used to extract more realistic tread images. The pressure images thereby are solely used to extract an image mask which can be iterated over the tire circumference images, as described in section 3.5, to extract multiple tread images for each tire. Due to the different scale and orientation of the standard images and the scans of the Fuji pressure foils, some processing steps are necessary. First, the tread pattern contours are extracted from the pressure foil scans with an active contour finder (from Scikit-Image [van14]). With the contour it is possible to extract the sole tire tread from the pressure foil scan. The tire tread is then projected on the standard image via template matching. This essentially ensures that the tread pattern scan is centered correctly on the standard image and scaling is done correctly. Through precalculating the scale of the mask through the known pixel representation in the tread image and the circumferential image, the proneness to errors in the template matching is reduced. The successful template matching leads to an image representation which has the same image height as the circumferential image, and width of the maximum width of the scaled tread pattern active contour. This image represents a mask that can be iterated over the circumferential image of each tire. Applying the same image processing steps as in previous described methods leads to images presented in fig. 4.51.

(a) Narrowest tire (205) **(b)** Mean tire (245) **(c)** Widest tire (285)
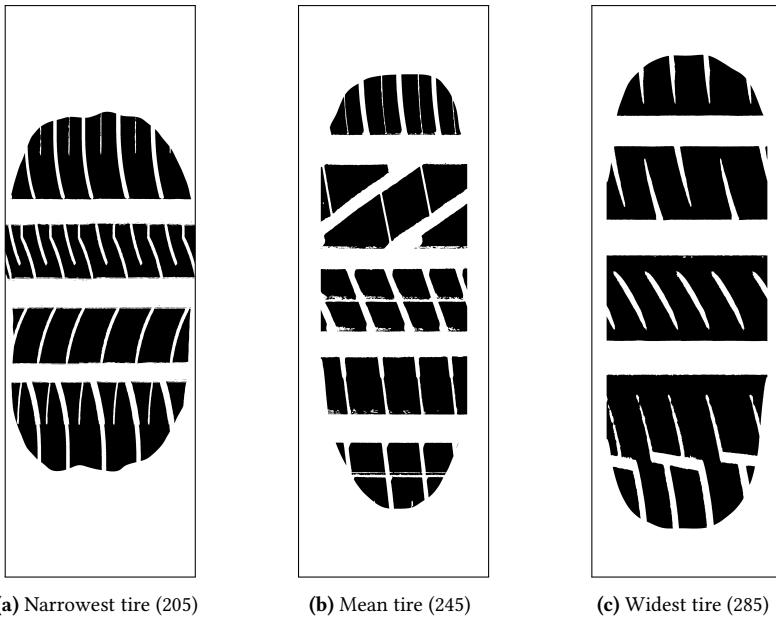
Figure 4.51: Three different tire images after cropping and rescaling to achieve a standardized format with each pixel size 1559 x 4707. Images extracted from tire circumference images with tread pattern shape achieved from Fuji Prescale® pressure foils. Frames only added for displaying purpose.

Table 4.12 displays the corresponding performance metric. For this observation not the whole data set of 42 tires is available anymore. Two tires (Tire 19 and 20 in table 3.1) were not available for the generation of the contact patches anymore, which is why these are not considered in this modeling approach. This reduces the data set to 40 tires and 1397 measurements.

Table 4.12: Performance metric for CMs for full test track with average value of both microphones and pressure foils as image input adaption calculated with LOOCV. All values in dB(A). RMSE values of third row are calculated for an output vector of dimension two.

Model	$\text{Mean}_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$\text{Mean}_{\text{RMSE, train}}$
$\text{CM}_{\text{Avg, Fuji}}^{\text{Distance}}$	0.58	0.15	0.98	0.32	0.37
$\text{CM}_{\text{Avg Fuji ,Test optimized}}^{\text{Distance}}$	0.51	0.13	0.98	0.32	0.59
$\text{CM}^{\text{Distance}}$	1.08	0.23	1.55	0.66	0.62

4.7 Combination of CNN and ANN over Distance for 1/3 Octave Levels

The prediction of the 1/3 octave SPL, requires additional input. The frequencies for 1/3 octave SPLs can be calculated with lower (f_1) and upper (f_2) as well as center frequency (f_0) as

$$f_1 = \frac{f_2}{\sqrt[3]{2}} \qquad f_2 = \sqrt[3]{2}f_1 \qquad f_0 = \sqrt[2]{f_1 * f_2} \qquad (4.5)$$

which results in 31 frequency bins in the range of 15.625 Hz to 16 000 Hz for f_0. The time series extracted from the tire models are transformed via a Fast Fourier Transformation (FFT) into a two sided frequency spectra with narrow, equidistant bins. These are then transformed into a two sided power density spectra. Generally, the prediction of spectral SPLs is even more useful than the prediction of overall SPLs. For tire manufacturers in particular, this provides insights into the various sound-generating mechanisms of the tires and thus enables targeted changes to the tire profile in order to weaken dominant mechanisms. By multiplying each frequency except 0 with two, a one-sides Power Spectral Density (PSD) is constructed [Cer00].

In the following, three different approaches for the additional frequency input are investigated. This input describes the 3D tire model through a spectral analysis and is therefore referred to as tire spectrum.

The first model $CM_{1/3\,\text{octave}}^{\text{PSD}}$ receives the same frequency bins as input as defined by the output. This, however, requires an adaptation of the initial frequency bins calculated through the FFT. The bins are defined by the sampling frequency, which is defined through the absolute sample point count and the time duration of the signal. However, the process described here, as well as in a previous publication [Leu23] does not guarantee identical frequency resolution in the FFT. The process is therefore adapted slightly to guarantee an identical sampling frequency for all tires. Since this leads to different length in the extracted time series, all series are zero padded to same size before applying the FFT.

In the first step, FFT signals are combined to the same frequency bins as defined through the output values. This is achieved by mapping each frequency of the PSD to its referring 1/3 octave frequency bin through their upper and lower frequencies. The PSD values of each bin are then added together to calculate the PSD value of the 1/3 octave frequency bins. This induced some error since frequency bandwidths of the PSD do not coincide with those of the 1/3 octave frequencies. However, this error is considered negligible and is performed equivalently on each tire. The frequency resolution of the original PSD was approximately 0.51 Hz.

In the second model $\mathrm{CM}_{\mathrm{Full}}^{\mathrm{PSD}}$, the frequencies are not combined into the 31 bins of the output. This is especially interesting since the combining of the frequencies loses some information. However, the maximum frequency of the input is defined as 4000 Hz and the frequency resolution is adapted to approximately 5.4 Hz. These two changes are applied to keep the number of input nodes referring to the PSD slightly lower. This leads to an overall number of 739 input values for each spectra.

The third model $\mathrm{CM}_{\mathrm{Embedded}}^{\mathrm{PSD}}$ minimizes the number of input nodes by using an embedding layer for the frequency spectra which combines the 739 different values into 31 values which are used as input for the fully connected ANN structure.

Fig. 4.52 shows the schematic sketch for the architecture and necessary input values abstracted for all approaches. The model predicts 31 1/3 octave levels for the left and right microphones, leading to 62 output variables for each prediction. Since this research always uses the sum of output variables in the RMSE calculation and not the mean of the difference in all output variables, these values are naturally larger than they might be in other literature.
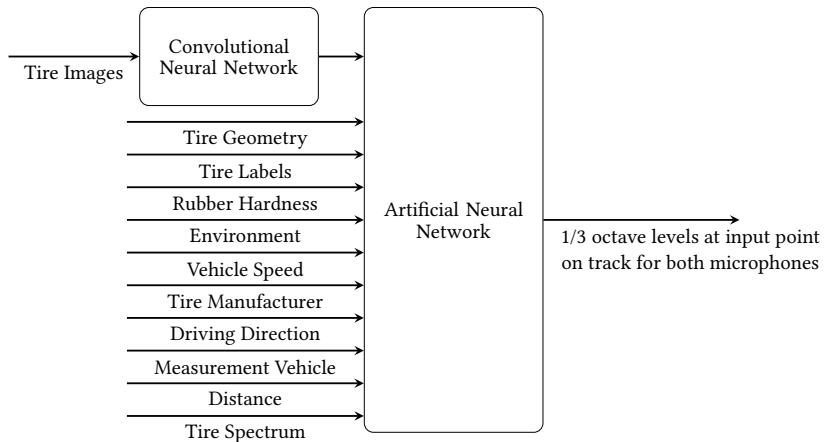
Figure 4.52: Principle sketch of combination of CM as distance model for 62 octave level output. Rubber hardness incorporates two hardness values with correlated surface temperature. Environment incorporates the wind speed and direction, air and road temperature, air pressure, and humidity. Tire geometry incorporates the tire nominal width, the aspect ratio, rim diameter, and profile depth. Tire labels refer to the load and speed index. Tire spectrum refers to the PSD extracted from digital tire models.

Table 4.13: Performance metric for CMs for full test track 1/3 octave level output calculated with LOOCV. All values in dB(A). The values are not averaged over the number of outputs, so that these values indicate the RMSE between vectors with 62 elements each.

Model	$\text{Mean}_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$\text{Mean}_{\text{RMSE, train}}$
$\text{CM}_{\text{1/3 octave}}^{\text{PSD}}$	24.83	3.71	33.19	18.73	13.72
$\text{CM}_{\text{Full}}^{\text{PSD}}$	23.86	3.72	32.29	18.18	11.19
$\text{CM}_{\text{Embedded}}^{\text{PSD}}$	23.31	3.60	30.22	17.89	13.75

For all approaches, the logarithm of the PSD values is used to push them in approximately the same scale as the 1/3 octave SPLs of the tires in dB(A). As shown in the LOOCV performance metric in table 4.13 the different approaches do not differ gravely in their performance. It is visible that the precalculated 1/3 octave levels lead to the largest mean RMSE and especially, the

minimum and maximum RMSE values are higher than in the other approaches. The approach to use all FFT values as input leads to the highest STD on the test set and lowest mean RMSE on the training data, which might indicate the greater possibility of overfitting due to the much larger input data and therefore overall more neurons since the networks are scaled according to the amount of input nodes. The third approach to use the FFT values in an embedding layer seems to output the best performing models. This is not only indicated by the overall lowest mean in RMSE but also the lowest STD indicating better generalization. This is further supported by the higher RMSE on the training data than without including an embedding layer.

The observation that embedding layers lead to the best performing models is also convenient computationally. The insertion of embedding layers lower the number of fully connected layers and with it the trainable parameters compared to the same input without embedding layers. This leads to lower computation time.

(a) Tire 0

(b) Tire 20

(c) Tire 40
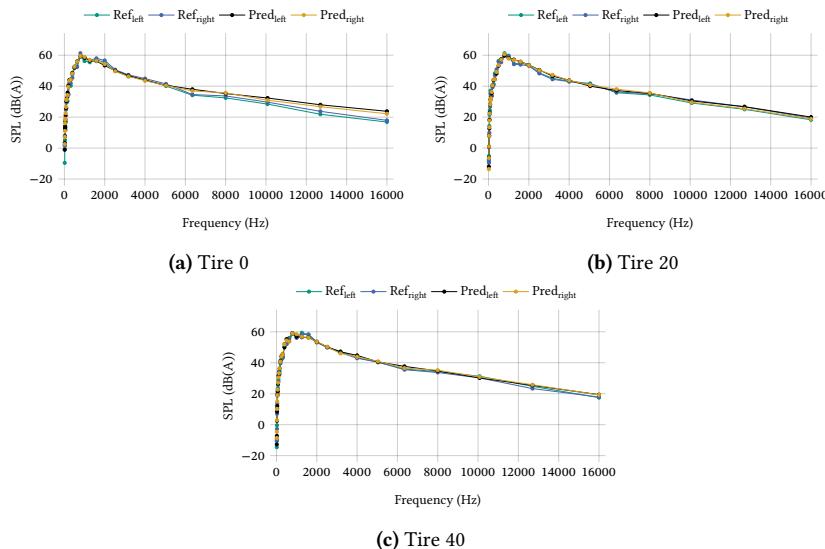
Figure 4.53: Example measurements of idle runs with $50\,\mathrm{km\,h^{-1}}$ at PP' for three different tires and their corresponding 1/3 octave SPLs for both microphones.

Fig. 4.53 displays the predictions and measurements of 1/3 octave SPLs for three different tires for a run at $50\,\mathrm{km\,h^{-1}}$ at position PP'. The predictions are made with the embedded model and each show the test tire that was left out according to LOOCV. Especially the peak between 750 Hz to 2000 Hz is predicted accurately for each tire. Fig. 4.54 shows the identical measurements and predictions as fig. 4.53 but for a restricted frequency range.

(a) Tire 0

(b) Tire 20

(c) Tire 40
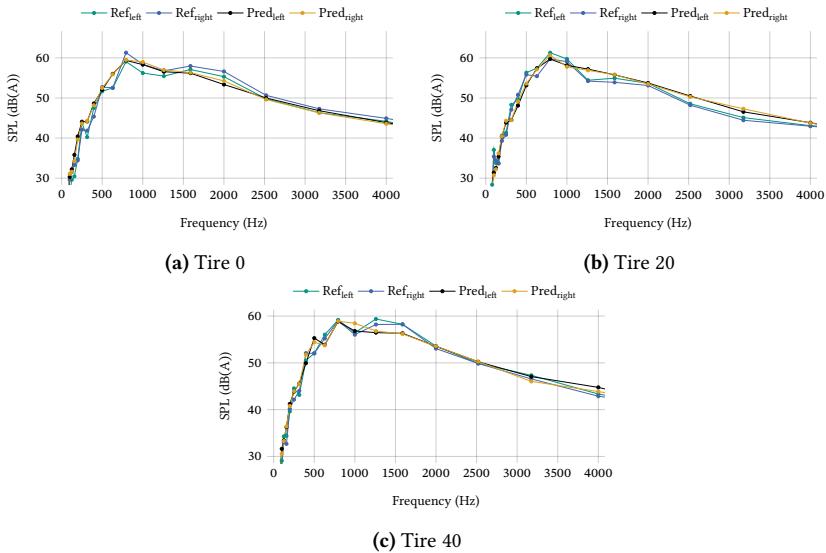
Figure 4.54: Example measurements of idle runs with $50\,\mathrm{km\,h^{-1}}$ at PP' for three different tires and their corresponding 1/3 octave SPLs for both microphones with limited axis range.

4.8 Bayesian Models

All previously calculated models are deterministic models in the sense that once trained, the output for identical inputs will always be identical. In contrast to probabilistic models, they do not offer implicit values of uncertainty other than those calculated through an evaluation metric. In the following,

two different probabilistic models which are implemented as Bayesian models in Pyro [Bin18], are presented. For the inference process, the No-U-Turn Sampler, which is an algorithm from the MCMC family introduced in [Hof14], is used.

Fig. 4.55 shows two different approaches of Bayesian models. The simpler model in fig. 4.55a refers to a Bayesian linear regression model which does not include an additional hierarchical structure. The more complex one in fig. 4.55b includes an additional hierarchical layer in the hyperparameter space to distinguish the tires explicitly in their TRN coefficients.

(a) Simple Model (Bayesian$_{\text{simple}}^{\text{PP'}}$)

(b) Complex Model (Bayesian$_{\text{complex}}^{\text{PP'}}$)
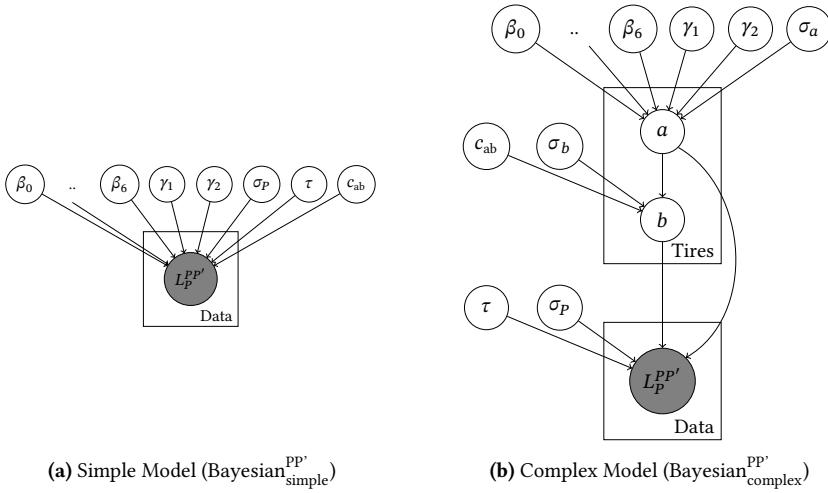
Figure 4.55: Plate notation for investigated Bayesian models.

The simpler model represents a model of the form

$$a = \beta_0 + \beta_1 c_{TW} + \beta_2 c_{AR} + \beta_3 c_{LI} + \beta_4 c_{RD}$$
$$+ \beta_5 (\gamma_1 - c_{GW})^2 + \beta_6 (\gamma_2 - c_{BG})^2 \tag{4.6}$$

$$b = c_{ab} - a \tag{4.7}$$

$$L_P^{PP'} = a + b \log(\frac{v}{v_{ref}}) + \tau \log(\frac{T_{air}}{T_{ref}}) \qquad . \tag{4.8}$$

where: | | |
|---|---|
| c_{ab} | Connector distribution of TRN coefficients initially sampled from $\mathcal{N}(42,5)$ |
| c_{AR} | Tire aspect ratio |
| c_{BG} | Tire block-to-groove ratio |
| c_{GW} | Average circumferential groove width |
| c_{LI} | Tire load index |
| c_{RD} | Rim diameter |
| c_{TW} | Tire nominal width |
| τ | Temperature coefficient initially sampled from $\mathcal{N}(2.6,0.1)$ |
| σ_a | STD of TRN coefficient a initially sampled from inverse Gamma distribution (3, 0.1) |
| σ_b | STD of TRN coefficient b initially sampled from inverse Gamma distribution (3, 0.1) |
| σ_P | STD of SPL initially sampled from inverse Gamma distribution (3, 0.1) |
| γ_1 | Center point of c_{GW} impact initially sampled from $\mathcal{U}(4,16)$ |
| γ_2 | Center point of c_{BG} impact initially sampled from $\mathcal{U}(0.1,0.9)$ |
| $\mathcal{U}(u_{low}, u_{high})$ | Uniform distribution with lower bound u_{low} and upper bound u_{high} |
| $\mathcal{N}(\mu_n, \sigma_n^2)$ | Gaussian distribution with mean μ_n and variance σ_n^2 |

The necessary block-to-groove ratio is calculated on basis of the pressure foils. Through the shape extraction necessary for tread pattern extraction, the block-to-groove ratio is calculated by dividing the included red pixels inside the

shape by the area enclosed by the extracted shape. The used priors in case for τ and c_{ab} can be considered informative since they include a mean value unequal to zero and compared to the mean the STD is not large. The priors for γ_1 and γ_2 can be considered weakly informative as they restrict the initial sample range but do not favor any particular values as these priors are uniformly distributed.

The variables a and b do not occur in the plate notation of the simpler Bayesian model since they represent fixed mathematical expressions and not distributions that are sampled from in the inference process. Eq. (4.8) is a combination of the earlier displayed speed relation between the SPL, the TRN (2.2), and the temperature correction (4.1). Eq. (4.7) enforces a relationship between the TRN coefficients a and b and eq. (4.6) builds a probabilistic regression model for the TRN in relation to tire parameters. In case of eq. (4.7) a linear relation would be more consistent with the observation in literature ([San02]) as well as the results in section 4.2.1. Yet, fig. 4.8 showed that the coefficients for the tires used in the presented research are Gaussian distributed so that a combination of the TRN coefficients with a distribution c_{ab} seems accurate enough. Also, combining the TRN coefficients with a linear equation adds much more complexity to the system since coefficient a is already an equation of multiple linear and quadratic terms. This leads to a multiple in calculation time, a loss in stability of the Markov chain, and no mentionable gain in the performance, which was analyzed in a smaller pre-study. The used parameters represent features that were recorded as important in earlier investigations in this dissertation as well as some in relation to the tire tread pattern based on the investigations of Ejsmont et al. [Ejs84]. During inference, the PDF of each latent variable displayed in fig. 4.55a is reconstructed to fit the PDF of the observed variable as close as possible.

In contrast to this, the second model, shown in fig. 4.55b, introduces an additional hierarchical layer. Tire coefficients a and b are no longer described as global distribution solely defined by the model and tire parameters. For each tire, a specific coefficient is defined. They are however incorporated from an underlying distribution which is defined similarly to eq. (4.6) and eq. (4.7). In this model, they represent latent variables which is why they appear in the

plate notation. They also receive priors for their assumed STD as indicated by σ_a and σ_b. c_{ab} represents the relation between a and b and the SPL is calculated equivalently to the simpler model.

However, the problem with using each tire as hierarchical instance is that it is not possible to make precise predictions on completely new tires. This is due to the fact that all tires share an underlying distribution for their TRN coefficients. However, these are conditioned in the inference process on the observed data. Simply forwarding the model therefore leads to much higher uncertainty. The more complex model is expected to represent the data much better than the simple model but incorporates difficulties in transferring the learned knowledge on completely unseen tires. This is especially seen in the fact that the STD in the TRN coefficients of each split is much higher on the test set compared to the training set. Fig. 4.56 demonstrates this spread in STD between test and train data.
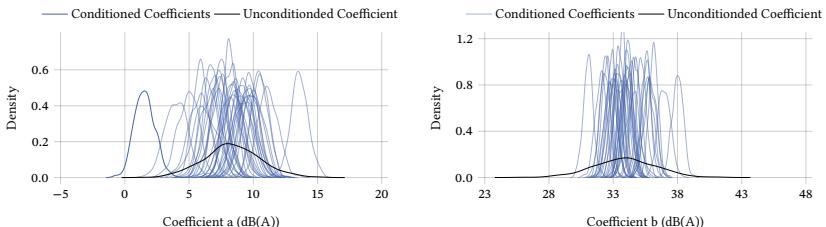
(a) PDF of a coefficients of hierarchical model for one specific data split to compare STD for tires with data to condition on and one tire without

(b) PDF of b coefficients of hierarchical model for one specific data split to compare STD for tires with data to condition on and one tire without

Figure 4.56: PDF for TRN coefficients in complex model for one specific data split.

To investigate the capability of the described models, 40 models for each of the plate notations are inferred. This refers to a LOOCV investigation with the data for which the pressure foil measurements are available. Table 4.14 shows the correlating performance metric. In case of the complex model, the inference process results in 40 TRN coefficients. The model does not receive data

to condition on for one tire so that always 1/40 tire coefficient combination is only sampled from the underlying parameter distributions.

Table 4.14: Performance metric for Bayesian regression models calculated with LOOCV. All values in dB(A).

Model	$Mean_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$Mean_{\text{RMSE, train}}$
$\text{Bayes}_{\text{simple}}^{\text{PP'}}$	0.97	0.60	2.79	0.33	1.02
$\text{Bayes}_{\text{complex}}^{\text{PP'}}$	1.16	0.64	2.88	0.28	0.28

In case of the simple model, the mean coefficients β_0 to β_4 do not change much for the different model splits. The mean values γ_1 and γ_2 are however rather volatile as seen in their PDF in fig. 4.57. Since all recorded block-to-groove ratios are between the values of 0.49 and 0.70, it is surprising that γ_2 changes its value mainly from 0.11 to 0.89, since this leads to completely opposing rankings in the block-to-groove rating of the tires. Furthermore, β_6 also changes its sign depending on γ_2, which makes it impossible to analyze a trend between the block-to-groove ratio and the SPL. Similar behavior is seen in case of γ_1 and β_5 only that the PDF of γ_1 is more widespread than the PDF of γ_2.

(a) PDF of γ_1 **(b)** PDF of γ_2
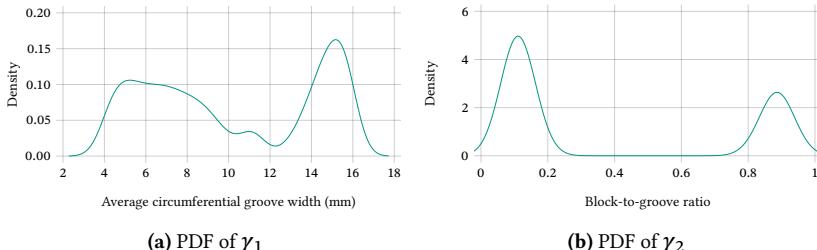
Figure 4.57: PDF for mean γ values for all model splits of the simple Bayesian model.

The overall spread in the model parameters and the thus occurring uncertainty in the predictions is best described through a PDF for the entire input data. Fig. 4.58 shows three representations of the simple model inferred on different data

splits with their performance on the left out test data. The first represents a model that performs well on test data, the second a model which performs not well on test data, and the last being a model trained on the whole data set to demonstrate the overall model capability to fit the data. It can be seen that the uncertainty of the model is rather large and does not vary much for one of the described models and data splits. This is not surprising since the model is rather simple and therefore does not offer many degrees of freedom to fit the data.

(a) PDF of test set with low RMSE

(b) PDF of test set with high RMSE

(c) PDF of data received from model inferred on all data
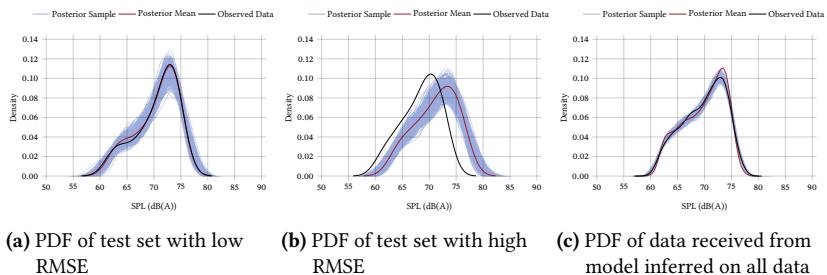
Figure 4.58: PDF for posterior samples, posterior mean sample, and observed data for the simple Bayesian model. The three figures refer to three models trained on different data. Figures a) and b) show PDF for one single tire left out for testing while figure c) shows the PDF for the entirety of data.

Even though it is reasonable to assume that the more complex model is less suitable to predict entirely new tires and tends to overfit the training data. This is also indicated by the higher maximum and mean RMSE values of the complex model compared to the simpler model as presented in table 4.14. This is a direct consequence of the lack of data to condition on and yet the additional degrees of freedom through describing each tire with individual TRN coefficients. The overlying distributions therefore have a high variance which leads to high variance in new predicted tires. This is demonstrated in the predicted TRN coefficients for the test tires shown in fig. 4.59. These have high variance and only differ slightly in their mean values achieved for different tires. This is especially different from the PDF for tires included in the training data, which reach much lower variance. This was shown for one model in fig. 4.56.
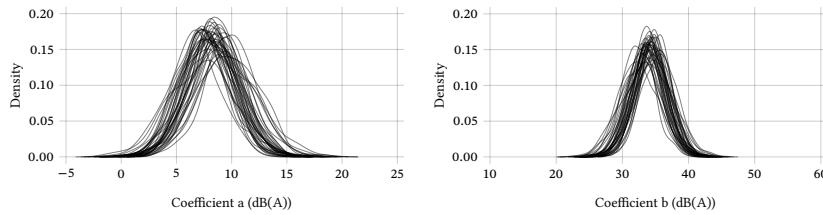
(a) PDF of a coefficients for hierarchical model of tires left out of training **(b)** PDF of b coefficients for hierarchical model of tires left out of training

Figure 4.59: PDF for TRN coefficients for all model splits for tire left out of training for complex Bayesian model.

The recorded mean values for γ_1 and γ_2 averaged over all data splits are with 11.08 mm and 0.50 more plausible than the values recorded for the simpler model. However, as seen in fig. 4.60 the complex model also results in much higher uncertainty of these parameters, which makes it questionable if the obtained values are reliable. This is especially the case for γ_2.
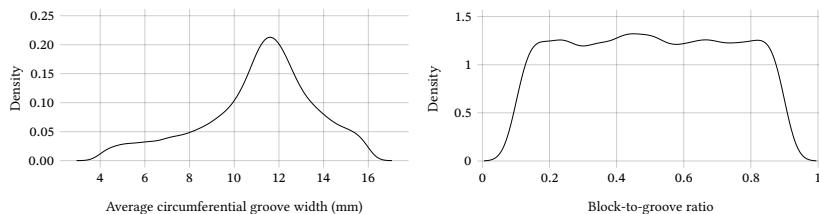
(a) PDF of γ_1 for all posterior samples of complex models **(b)** PDF of γ_2 for all posterior samples of complex models

Figure 4.60: PDF for γ values for all model splits of the complex Bayesian model calculated for the entirety of posterior samples.

The investigation of the general data fit of the complex model is shown in fig. 4.61. The results are distinct from those of the simpler model. The model performance on test data sets for a tire that is predicted with low and one with high RMSE, shown in fig. 4.61a and fig. 4.61b, both have exceptionally high variance. In case of the model which received the entire data as training data, the data fit is exceptionally good. This, however, is not surprising since the

model infers TRN coefficients for each tire only with the condition of them being part of an underlying distribution.

(a) PDF of test set with low RMSE

(b) PDF of test set with high RMSE

(c) PDF of data received from model inferred on all data
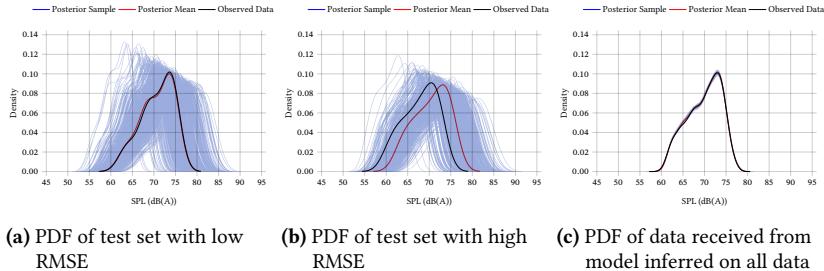
Figure 4.61: PDF for posterior samples, posterior mean sample, and observed data for the complex Bayesian model. The three figures refer to three models trained on different data. Figures a) and b) show PDF for one single tire left out for testing while figure c) shows the PDF for the entirety of data.

It can be summarized that both Bayesian models perform worse on the test set than the previously considered $CM^{PP'}$. As shown in table 4.14, this is especially the case for the maximum RMSE, the mean RMSE is only marginally smaller. This might result from the highly non-linear interaction of the relevant features, which the Bayesian models as considered here are not capable to reproduce. Still the Bayesian model offer easier interpretation and do not require as much data as the DL models for training, which is why they can still be considered useful. Also, it can be stated that the complex Bayesian model fits the training data even better than the $CM^{PP'}$. This indicates overfitting, so that an adjustment of the mathematical representations in the complex Bayesian model should be considered.

5   Conclusion and Summary

This research investigates different models on their suitability to predict the Tire Road Noise (TRN) on an ISO 10844 test track. This is done on the basis of 3D tire models reconstructed through photogrammetry, pressure foil measurements, and Coast-by (CB) measurements. Throughout this dissertation, the complexity of the models is continuously increased. The features used in different models include the vehicle speed, the position of the vehicle on the test track, the tire nominal width, aspect ratio, rim diameter, shore hardness with measurement temperature, load index, speed index, tire manufacturer, measurement vehicle, driving direction, air temperature, road temperature, air pressure, air humidity, wind speed, wind direction, tread block-to-groove ratio, average width of circumferential grooves, tire tread pattern spectra, the total tread pattern design as well as the shape of the contact patch.

In this dissertation, the Root Mean Squared Error (RMSE) values are calculated as sum of the output variables which naturally leads to larger errors for more output variables of a model. This was decided since it indicates the overall performance in output values, however in literature it is more common to average over the output variables. Therefore, the values in this chapter are adapted to the more common form in literature, to make cross comparison easier.

The simplest models (section 4.2) predict the TRN at one point of the measurement track for the average of both microphones and uses linear and Least Absolute Shrinkage and Selection Operation (lasso) regression. Even though their mean RMSE values obtained through Leave-One-Out Cross-Validation (LOOCV) could be considered good (1.11 dB(A), 1.09 dB(A), and 1.03 dB(A)) for the simplicity of the models, the maximum recorded RMSEs are high (2.54 dB(A), 2.15 dB(A), and 2.08 dB(A)), which indicates that the models do

not generalize well. Still, the lasso regression in particular provides important information, as it can also be regarded as a method for feature selection. For the chosen 17 features, lasso only results in the vehicle speed, the tire nominal width, the air pressure, the road temperature, and the load index having acknowledgeable impact on the prediction.

To incorporate nonlinearities, Artificial Neural Networks (ANNs) are also investigated (section 4.3). For this, features in nominal scale are one-hot encoded. These, as well as the shore hardness values with correlating surface temperature measurements, are preprocessed in an embedding layer before being passed to the ANN. The performance of the ANN for the two independent microphones at one point of the test track is much better than in case of the linear and lasso regression. This includes the mean (0.71 dB(A)) as well as maximum RMSE (1.30 dB(A)). Throughout the model design process, it is shown that hyperparameter tuning does not lead to acknowledgeable better performance of the models. Yet, an adaptive learning rate possibly enables a broader coverage of the parameter space and at the same time allows small steps in the last iterations.

In contrast to the ANN, which does not include information of the tire tread, except the tire nominal width, Convolutional Neural Networks (CNNs) (section 4.4) are constructed which receive tire tread images of each tire. These images are extracted from the 3D models of the tires and scaled so that the representation in millimeters per pixel is identical for each image. These images only include a fragment of the tire, so that it is possible to extract multiple images from one tire which are thus not identical. This allows for a much bigger data set for training. For the CNN, only measurements in the range of $78.5\,\mathrm{km\,h^{-1}}$ to $81.5\,\mathrm{km\,h^{-1}}$ are included. Hyperparameter optimization once again does not result in much better performance of the model. However, a rather small architecture with only three convolutional layers and three maximum pooling layers shows satisfying performance on the mean RMSE (0.52 dB(A)). The problem is seen in the high maximum RMSE (1.48 dB(A)), which is 0.18 dB(A) larger than the one recorded for the ANN. In case of the CNN, the model performance is strongly impacted by the left out tire, which is also a result of the imbalance of the data. Model interpretation with Shapley

Additive Explanations (SHAP) values on pixel basis did not show any noticeable trends, hereby it seemed that the largest influence comes from the shape of the tread pattern passed to the CNN. Still, the CNN provides valuable information and especially shows, that the model is capable of learning to distinguish different tires from only a different fragment of the same tire. Also, this approach shows that the Sound Pressure Level (SPL) of a tire is determined through the tread pattern.

In the next step, the ANN and CNN are combined (section 4.5) into a multimodal model that allows a simultaneous input of the tire tread image and the numerical values. These describe the tire, operating conditions, and environment to predict the SPL on one point of the track for both microphones. The thus produced model performs better than the ANN for all metrics under consideration. Especially, the maximum RMSE is 0.18 dB(A) lower. The combination of many models that are trained each for a specific point on the test track to predict the SPL over the test track also reaches satisfying results. Additionally including the location on the test track as feature in the model leads to much steadier predictions of SPL curves. The investigation of some models that are trained with certain features left out to decrease multicollinearity show the following trends. The vehicle speed is the most important feature, followed by the location on the test track, and the tire nominal width. These are all positively correlated with the SPL. The air temperature and the tire load index have a strong negative correlation with the SPL. These results are obtained through applying Integrated Gradients (IG) as well as manually investigating the model prediction through interactively changing single features in a representative batch of samples. The model also predicts the overall shape of the SPL curves for most tires well. This could indicate that the model learns the radiation characteristic of the tires up to a certain degree and with this indicates that the average tire tends to stronger radiation from the leading edge than the trailing edge, which could also be seen for the average measurements.

In addition to the prediction of two independent SPL for both microphones, model performance for 1/3 octave levels is also part of this research (section 4.7). This is possible by providing an additional tread pattern spectrum as input to the model to incorporate knowledge about the randomization and

offset. These spectra are obtained through virtually rolling the 3D tire models on a plane road and using the thus achieved contact pattern as input for an FFT. This research investigates three different processes of passing these spectra to the model. As raw output of the FFT directly to the model, as raw output of the FFT to an embedding layer, and as preprocessed 1/3 octave levels. Even though the models do not vary gravely (mean RMSE of 3.15 dB(A), 3.03 dB(A), and 2.96 dB(A) for 1/3 octave levels, FFT, and embedding input) in their performance, the embedding layer performs slightly better on the other metrics calculated for the test data set. It can be noted that all models approximate the measured 1/3 octave levels well in the range of investigation of 15.625 Hz to 16 000.0 Hz. Especially the peak at around 750 Hz to 2000 Hz, which is also the range expected from literature, is well approximated.

Since the measurements show some unsteady behavior and discrepancy between the measured SPLs for both microphones, a last DL iteration is performed which receives the average SPL of both microphones as labels for training (section 4.6.3). This model is better than any of the previous models. It reaches a mean RMSE of 0.64 dB(A) with maximum 0.93 dB(A) for the prediction of the whole measurement track when the model that minimizes the combined loss of test and training set is considered. In case of the model that performs best on the test set, the values are even as low as 0.60 dB(A) for the mean RMSE and a maximum RMSE of 0.91 dB(A) while the mean training RMSE only increases by 0.11 dB(A). As a variation of this model, the extracted tread pattern images from pressure foils are used as image input. This results in only slightly different values with mean RMSE of 0.58 dB(A) and maximum 0.98 dB(A) for the Test-Train (TT) models and a mean RMSE of 0.51 dB(A) and maximum 0.98 dB(A) for the models performing best on the test data. In this case only 40 of the previous 42 tires could be used since only for these pressure foils are available. Still, this shows that the earlier approach of modeling the tire tread pattern as simple rectangle with dimensions linked to these of the tires, also leads to high model accuracy. These models can still be improved, but the model which receives a rectangular tread pattern seems to offers the best trade-off between model accuracy and data acquisition effort.

In contrast to the linear models, ANNs, and CNNs, two Bayesian models are also investigated (section 4.8). One embodies a rather simple Bayesian linear regression while the other includes an additional hierarchical layer to allow model pooling. Both models are based on the speed SPL relation of the tire. The easier model assumes a global probability distribution of these parameters and the model with additional hierarchical layer infers one set of TRN coefficients for each tire, and extracts these from an underlying density function, which connects all tires. Even though especially the hierarchical model is capable of representing the distribution of the data well, the performance on the test sets, especially the maximum RMSE, is with 2.79 dB(A) and 2.88 dB(A) for the simpler and the more complex model much higher than the CM for position PP', which results in 1.12 dB(A). The Bayesian models are trained with the tire nominal width, tire aspect ratio, rim diameter, load index, air temperature, average circumferential groove width, and block-to-groove ratio extracted from the pressure foil measurements. In case of the block-to-groove ratio, no clear trend could be extracted. The average groove width however, showed to result in the lowest SPL for 11.08 mm.

6 Outlook

While this dissertation shows the capability of purely empirically predicting CB measurements, the model performance can still be improved in the future. This might be achieved through different approaches. While the data used for this research was restricted in some parameters, such as the wind speed, the deviation between runs and or between the vehicle sides was not used to cleanse the data set. Eliminating measurements that differ stronger than expected could lead to a better data set and with it more accurate predictions. However, especially the difference in recorded SPL for both sides of the tires could also be used to understand the acoustic process even better and abbreviate design criteria for less noisy tires. To add onto this, even larger measurement campaigns could be performed to enlarge the overall data set and also ensure that the data are balanced. E.g., same amount of tire fragments as image input and measurements per tire. This could also include many more measurement vehicles so that future models might gain advantage through including vehicle distinct features as length, mass, drag coefficient, ground clearance, wheelhouse geometry, and absorption characteristic of the vehicle underbody. The biggest gain however could be reached by simply adding more tires to the data set which are representative for a common target market.

Also, the empirical models could be connected with more physical sub models. Especially a precalculation of the horn effect for each of the independent tires might lead to an even better representation of the tire radiation characteristic and thus increase model accuracy. Furthermore, the images extracted from the 3D tire models constructed with photogrammetric methods also induce some error into the model. This results from digitization errors which extend to some artifacts in the images, some additional grooves, or some grooves to be closed. Improving this process therefore also enables higher model accuracy.

Even though the models described in this dissertation are capable to predict each point on the test track and also learn the relation between these points, it might still be interesting to have a full CB measurement predicted at once and not only by parts. Ongoing research could also include hyperparameter optimizations with different sampling in hyperparameter space. This should include deeper, pretrained models or transformers.

Through larger calculation power, it might also be possible to design models that receive the actual 3D tire models and not only fragments of the tread pattern as images. These could also be in the shape of the tire silhouette presented in this research.

In terms of model usability, the presented approach could be extended to multiple measurement tracks. An inclusion of pavement characteristics such as micro- and macro texture, absorption coefficient might even enable the transfer from one measurement track to another. For this, especially the Bayesian model with additional hierarchical structure could be extraordinarily suitable. Using each measurement track as component in the hierarchical structure allows linking the test tracks but keep a certain degree of freedom in the inferred parameters. Adding surface parameters can generally be expected to lead to higher model accuracy since the interaction of tire and pavement is highly nonlinear and defined by the surface pairing.

Also, this research indicated that the load index of tires plays a crucial role for the emitted exterior noise. Since this is nowhere else recorded in literature, it should be investigated further. It might especially be reasonable to take a look into how the load index is obtained by tire manufacturers and use the thereby necessary parameters as features in similar models as shown in this dissertation. This is especially interesting since it could show if the tire load index indeed has a large effect on the TRN or if it merely results from the high correlation of the tire load index and the outer tire diameter.

Lastly, since this research focused on the free rolling tire but for certification acceleration is also necessary, this should be a main focus in further research. It is generally known that a positive correlation between the drive torque

applied on tires and the SPL exists. However, the tire parameters explicitly influencing this behavior are not well researched yet.

Bibliography

[Abr05]   ABRAHAM, Ajith: "Artificial neural networks". In: *Handbook of measuring system design*. Chichester: Wiley, 2005, pp. 901–903. ISBN: 0-470-02143-8 (cit. on pp. 29, 33).

[Ada18]   ADADI, Amina and BERRADA, Mohammed: "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)". In: *IEEE Access* 6 (2018), pp. 52138–52160. DOI: 10.1109/ACCESS.2018.2870052 (cit. on pp. 41, 42).

[Ahs21]   AHSAN, Md; MAHMUD, M.; SAHA, Pritom; GUPTA, Kishor and SIDDIQUE, Zahed: "Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance". In: *Technologies* 9.3 (2021), p. 52. DOI: 10.3390/technologies9030052 (cit. on p. 54).

[Aki15]   AKINWANDE, Michael Olusegun; DIKKO, Hussaini Garba and SAMSON, Agboola: "Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis". In: *Open Journal of Statistics* 05.07 (2015), pp. 754–767. ISSN: 2161-718X. DOI: 10.4236/ojs.2015.57075 (cit. on p. 27).

[Aki19]   AKIBA, Takuya; SANO, Shotaro; YANASE, Toshihiko; OHTA, Takeru and KOYAMA, Masanori: "Optuna: A Next-generation Hyperparameter Optimization Framework". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019 (cit. on p. 84).

[Al 06]   AL SHALABI, Luai and SHAABAN, Zyad: "Normalization as a preprocessing engine for data mining and the approach of preference matrix". In: *2006 International conference on dependability of computer systems*. 2006, pp. 207–214 (cit. on p. 54).

[Alo19]   Alom, Md Zahangir; Taha, Tarek M.; Yakopcic, Chris; West-
           berg, Stefan; Sidike, Paheding; Nasrin, Mst Shamima; Hasan,
           Mahmudul; van Essen, Brian C.; Awwal, Abdul A. S. and
           Asari, Vijayan K.: "A State-of-the-Art Survey on Deep Learning
           Theory and Architectures". In: *Electronics* 8.3 (2019), p. 292. DOI:
           10.3390/electronics8030292 (cit. on p. 36).

[Ame16]   American Association of State Highway and Transporta-
           tion Officials: Standard Method of Test for Measurement
           of Tire/Pavement Noise Using the On-Board Sound Intensity
           (OBSI) Method. Washington D., United States of America, 2016
           (cit. on p. 4).

[And06]   Andersen, Bent; Kragh, Jørgen and Bendtsen, Hans: Acoustic
           performance: Low noise road pavements. Vol. 44. VI eksternt
           notat. Hedehusene, 2006. ISBN: 87-91177-94-4 (cit. on p. 17).

[Anf07]   Anfosso-Lédée, Fabienne and Pichaud, Yves: "Temperature ef-
           fect on tyre–road noise". In: *Applied Acoustics* 68.1 (2007), pp. 1–
           16. ISSN: 0003682X. DOI: 10.1016/j.apacoust.2006.06.001 (cit. on
           pp. 17, 18).

[Bar12]   Barber, David: Bayesian Reasoning and Machine Learning.
           Cambridge University Press, 2012. ISBN: 978-0-521-51814-7 (cit.
           on p. 37).

[Bar22]   Barrigón Morillas, Juan Miguel; Rey Gozalo, Guillermo;
           Montes González, David; Sánchez-Fernández, Manuel and
           Bachiller León, Alicia: "A comprehensive experimental study
           of the influence of temperature on urban road traffic noise un-
           der real-world conditions". In: *Environmental pollution (Barking,
           Essex : 1987)* 309 (2022), p. 119761. DOI: 10.1016/j.envpol.2022.
           119761 (cit. on p. 21).

[Bec08]   Beckenbauer, Thomas; Klein, Philippe; Hamet, Jean–François
           and Kropp, Wolfgang: "Tyre/road noise prediction: A compari-
           son between the SPERoN and HyRoNE models – Part 1". In: *The
           Journal of the Acoustical Society of America* 123.5 (2008), p. 3388.
           ISSN: 0001-4966. DOI: 10.1121/1.2934048 (cit. on pp. 21, 22).

[Bec12]   BECKENBAUER, Thomas; ANGST, Christian and BUERGISSER, Philip: "Simulation of tyre/road noise as a tool for the evaluation of the acoustic behaviour of road surfaces". In: *Eurasphalt & Eurobitume Congress* 5 (2012), p. 338 (cit. on p. 22).

[Bel99]   BELL, Tyler; LI, Beiwen and ZHANG, Song: "Structured light techniques and applications". In: *Wiley Encyclopedia of Electrical and Electronics Engineering* (1999), pp. 1–24 (cit. on pp. 55, 56).

[Ben09]   BENDTSEN, Hans; KOHLER, Erwin and QING, Lu: Temperature influence on road traffic noise: Californian OBSI measurement study. Vol. 169. VI rapport. Vejdirektoratet, 2009. ISBN: 978-87-92094-47-6 (cit. on p. 18).

[Ber19]   BERRAR, Daniel: "Cross-Validation". In: *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier, 2019, pp. 542–545. ISBN: 9780128114322. DOI: [10.1016/B978-0-12-809633-8.20349-X](https://doi.org/10.1016/B978-0-12-809633-8.20349-X) (cit. on p. 40).

[Bin18]   BINGHAM, Eli; CHEN, Jonathan P.; JANKOWIAK, Martin; OBERMEYER, Fritz; PRADHAN, Neeraj; KARALETSOS, Theofanis; SINGH, Rohit; SZERLIP, Paul; HORSFALL, Paul and GOODMAN, Noah D.: "Pyro: Deep Universal Probabilistic Programming". In: *Journal of Machine Learning Research* (2018) (cit. on p. 136).

[Boy18]   BOYD, Stephen P. and VANDENBERGHE, Lieven: Introduction to applied linear algebra: Vectors, matrices, and least squares. Cambridge, UK and New York, USA: Cambridge University Press, 2018. ISBN: 978-1-316-51896-0 (cit. on p. 26).

[Bri23]   BRIDGESTONE AMERICAS, INC.: Tire Terminology. 2023. URL: https://www.bridgestoneamericas.com/en/company/safety/tires-101/tire-terminology# (visited on 11/19/2023) (cit. on p. 9).

[Bro11]   BROOKS, Steve; GELMAN, Andrew; JONES, Galin L. and MENG, Xiao-Li, eds.: Handbook of Markov chain Monte Carlo. Chapman & Hall/CRC handbooks of modern statistical methods. Boca Raton, Fla.: CRC Press Taylor & Francis, 2011. ISBN:

1420079417. URL: https://www.mcmchandbook.net/index.html (cit. on p. 38).

[Büh11]   BÜHLMANN, Erik and ZIEGLER, Toni: "Temperature effects on tyre/road noise measurements". In: *Proceedings of Inter-Noise 2011*. 2011, pp. 4–7 (cit. on pp. 18, 19).

[Büh13]   BÜHLMANN, Erik; EGGER, Sebastian and ZIEGLER, Toni: "Ageing of the new CPX reference tyres during a measurement season". In: *Proceedings of Inter-Noise 2013*. 2013 (cit. on p. 9).

[Büh14]   BÜHLMANN, Erik and VAN BLOKLAND, Gijsjan: "Temperature effects on tyre/road-noise–A review of empirical research". In: *Proceedings of Forum Acusticum*. 2014, p. 6 (cit. on p. 19).

[Büh18]   BÜHLMANN, Erik; EGGER, Sebastian; MIODUSZEWSKI, Piotr and SANDBERG, Ulf: "An in-depth look at the tire rubber hardness influence on tire/road noise measurements". In: *Proceedings of Inter-Noise 2018 conference 2018*. 2018 (cit. on pp. 8, 9).

[Car23]   CARL ZEISS GOM METROLOGY GMBH: Triangulation: bewährtes Messprinzip mit Zukunft. 2023. URL: https://www.gom.com/de-de/themen/triangulation (visited on 01/06/2023) (cit. on pp. 55, 56).

[Cer00]   CERNA, Michael and HARVEY, Audrey F.: The fundamentals of FFT-based signal analysis and measurement. 2000 (cit. on p. 131).

[Cha07]   CHANDRASEKARAN, V. C.: Essential rubber formulary: Formulas for practitioners. PDL, Plastics Design Library. Norwich, NY: William Andrew Publ, 2007. ISBN: 978-0-8155-1539-5. URL: http://www.loc.gov/catdir/enhancements/fy0905/2006103071-d.html (cit. on p. 7).

[Che21]   CHEN, Liuxiao; CONG, Lin; DONG, Yuanshuai; YANG, Guang; TANG, Boming; WANG, Xudong and GONG, Hongren: "Investigation of influential factors of tire/pavement noise: A multilevel

Bayesian analysis of full-scale track testing data". In: *Construction and Building Materials* 270 (2021), p. 121484. ISSN: 09500618. DOI: [10.1016/j.conbuildmat.2020.121484](cit. on pp. 14, 23).

[Cle15]   CLEVERT, Djork-Arné; UNTERTHINER, Thomas and HOCHREITER, Sepp: Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). 2015. URL: http://arxiv.org/pdf/1511.07289v5 (visited on 01/25/2024) (cit. on p. 104).

[Dar68]   DARLINGTON, Richard B.: "Multiple regression in psychological research and practice". In: *Psychological bulletin* 69.3 (1968), p. 161 (cit. on p. 46).

[Das20]   DAS, Arun and RAD, Paul: Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. 2020. DOI: [10.48550/arXiv.2006.11371](cit. on p. 42).

[Deb02]   DEB, K.; PRATAP, A.; AGARWAL, S. and MEYARIVAN, T.: "A fast and elitist multiobjective genetic algorithm: NSGA-II". In: *IEEE Transactions on Evolutionary Computation* 6.2 (2002), pp. 182–197. DOI: [10.1109/4235.996017](cit. on p. 84).

[Deu03]   DEUTSCHES INSTITUT FÜR NORMUNG: Kunststoffe und Hartgummi - Bestimmung der Eindruckhärte mit einem Durometer (Shore-Härte) (ISO 868:2003); Deutsche Fassung EN ISO 868:2003. Berlin, 2003. DOI: [10.31030/9413329](cit. on p. 7).

[Deu17]   DEUTSCHES INSTITUT FÜR NORMUNG: Akustik - Messung des Einflusses von Straßenoberflächen auf Verkehrsgeräusche - Teil 2: Nahfeldmessverfahren (ISO 11819-2:2017); Deutsche Fassung EN ISO 11819-2:2017. Berlin, 2017. DOI: [10.31030/2575051](cit. on p. 4).

[Deu22]   DEUTSCHES INSTITUT FÜR NORMUNG: Akustik - Messung des Einflusses von Straßenoberflächen auf Verkehrsgeräusche. Teil 1, Statistisches Vorbeifahrtverfahren (ISO/DIS 11819-1:2021); Deutsche und Englische Fassung prEN ISO 118-1:2021. Berlin, 2022 (cit. on p. 4).

[Deu23]   Deutsches Institut für Normung: DIN ISO 10844, Akustik - Anforderungen an Prüfstrecken zur Messung der Geräuschemission von Straßenfahrzeugen und ihren Reifen (ISO 10844:2021). Berlin, 2023 (cit. on p. 49).

[Don21]   Dong, Shi; Wang, Ping and Abbas, Khushnood: "A survey on deep learning and its applications". In: *Computer Science Review* 40 (2021), p. 100379. issn: 15740137. doi: 10.1016/j.cosrev.2021.100379 (cit. on p. 36).

[Ejs82]   Ejsmont, Jerzy A.: Tire/road noise: comparison of road and laboratory measurements and influence of some tire parameters on generation of sound. Statens Väg-och Trafikinstitut. VTI Rapport nr 244A, 1982 (cit. on pp. 6, 7, 14).

[Ejs84]   Ejsmont, Jerzy A.; Sandberg, Ulf and Taryma, Stanislaw: "Influence of tread pattern on tire/road noise". In: *SAE Transactions* (1984), pp. 632–640 (cit. on pp. 10, 11, 138).

[Eki15]   Ekinci, Ş.; Çarman, K.; Taşyürek, M. and Mirik, M.: "Relationship between Deflection and Contact Area of Drive Tire". In: *International Journal of Materials, Mechanics and Manufacturing* 4.3 (2015), pp. 179–182. issn: 17938198. doi: 10.7763/IJMMM.2016.V4.251 (cit. on p. 15).

[Eur20]   European Environment Agency: Exposure of Europe's population to environmental noise. Vol. 22/2019. EEA report. Copenhagen: EEA, 2020. isbn: 978-92-9480-209-5. url: https://www.eea.europa.eu/publications/environmental-noise-in-europe/ (visited on 12/14/2023) (cit. on p. 1).

[Feu19]   Feurer, Matthias and Hutter, Frank: "Hyperparameter Optimization". In: *Automated Machine Learning: Methods, Systems, Challenges*. Ed. by Hutter, Frank; Kotthoff, Lars and Vanschoren, Joaquin. Cham: Springer International Publishing, 2019, pp. 3–33. isbn: 978-3-030-05318-5. doi: 10.1007/978-3-030-05318-5_1 (cit. on p. 41).

[For18]   Forssén, Jens; Hoffmann, Alice and Kropp, Wolfgang: "Auralization model for the perceptual evaluation of tyre–road noise". In: *Applied Acoustics* 132 (2018), pp. 232–240. issn: 0003682X. doi: 10.1016/j.apacoust.2017.11.023 (cit. on p. 22).

[Fra57]   Frank, Rosenblatt: "The perceptron, a perceiving and recognizing automaton Project Para". In: *Cornell Aeronautical Laboratory* (1957) (cit. on p. 29).

[Fuj09]   Fujiwara, S.; Yumii, K.; Saguchi, T. and Kato, K.: "Reduction of Tire Groove Noise Using Slot Resonators". In: *Tire Science and Technology* 37.3 (2009), pp. 207–223. issn: 0090-8657. doi: 10.2346/1.3138766 (cit. on p. 12).

[Fuj11]   Fujifilm Prescale: Tactile pressure indicating sensor film. 2011. url: https://www.sensorprod.com/prescale/application-pages/tire-tread/fuji-prescale-tire-tread.pdf (visited on 12/13/2023) (cit. on pp. 57, 58).

[FUJ23]   FUJIFILM Europe GmbH: Prescale. 2023. url: https://www.fujifilm.com/de/de/business/inspection/measurement-film/prescale/feature (visited on 12/13/2023) (cit. on p. 58).

[Gei20]   Geirhos, Robert; Jacobsen, Jörn-Henrik; Michaelis, Claudio; Zemel, Richard; Brendel, Wieland; Bethge, Matthias and Wichmann, Felix A.: "Shortcut learning in deep neural networks". In: *Nature Machine Intelligence* 2.11 (2020), pp. 665–673. doi: 10.1038/s42256-020-00257-z (cit. on p. 41).

[Gel13]   Gelman, Andrew; Carlin, John B.; Stern, Hal S.; Dunson, David B.; Vehtari, Aki and Rubin, Donald B.: Bayesian Data Analysis. Chapman and Hall/CRC, 2013. isbn: 9780429113079. doi: 10.1201/b16018 (cit. on pp. 38, 39).

[Gel20]   Gelman, Andrew; Vehtari, Aki; Simpson, Daniel; Margossian, Charles C.; Carpenter, Bob; Yao, Yuling; Kennedy, Lauren; Gabry, Jonah; Bürkner, Paul-Christian and Modrák, Martin: Bayesian Workflow. 2020. url: http://arxiv.org/pdf/2011.01808v1 (visited on 01/25/2024) (cit. on p. 39).

[Gha15]   Ghahramani, Zoubin: "Probabilistic machine learning and arti-
           ficial intelligence". In: *Nature* 521.7553 (2015), pp. 452–459. doi:
           [10.1038/nature14541](#) (cit. on p. 37).

[Gui19]   Guidotti, Riccardo; Monreale, Anna; Ruggieri, Salvatore;
           Turini, Franco; Giannotti, Fosca and Pedreschi, Dino: "A
           Survey of Methods for Explaining Black Box Models". In: *ACM
           Computing Surveys* 51.5 (2019), pp. 1–42. issn: 0360-0300. doi:
           [10.1145/3236009](#) (cit. on pp. 41, 42).

[Guo16]   Guo, Cheng and Berkhahn, Felix: Entity Embeddings of Cate-
           gorical Variables. 2016. url: http://arxiv.org/pdf/1604.06737v1
           (visited on 01/25/2024) (cit. on p. 83).

[Han11]   Han, Jiawei; Kamber, Micheline and Pei, Jian: Data mining:
           Concepts and techniques. 3rd ed (Online-Ausg.) Amsterdam,
           Boston and Heidelberg: Elsevier Morgan Kaufmann, 2011. isbn:
           978-0-12-381479-1. url: [https://ebookcentral.proquest.com/lib/
           kxp/detail.action?docID=729031](https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=729031) (cit. on pp. 54, 86).

[Haw04]   Hawkins, Douglas M.: "The problem of overfitting". In: *Journal
           of chemical information and computer sciences* 44.1 (2004), pp. 1–
           12. issn: 0095-2338. doi: [10.1021/ci0342472](#) (cit. on p. 27).

[Hea94]   Health Council of the Netherlands: Noise and health:
           Report by a committee of the Health Council of the Nether-
           lands. Vol. 1994,15E. Publication / Gezondheidsraad. The Hague:
           Health Council of the Netherlands, 1994. isbn: 90-5549-048-2.
           url: [https://www.healthcouncil.nl/documents/advisory-
           reports/1994/09/15/noise-and-health](https://www.healthcouncil.nl/documents/advisory-reports/1994/09/15/noise-and-health) (visited on 12/14/2023)
           (cit. on p. 1).

[Heh20]   Hehenberger, Peter: Computerunterstützte Produktion. Berlin,
           Heidelberg: Springer Berlin Heidelberg, 2020. isbn: 978-3-662-
           60875-3. doi: [10.1007/978-3-662-60876-0](#) (cit. on p. 56).

[Ho13]   Ho, Ka-Yee; Hung, Wing-Tat; Ng, Chung-Fai; Lam, Yat-Ken;
           Leung, Randolph and Kam, Eddy: "The effects of road surface
           and tyre deterioration on tyre/road noise emission". In: *Applied*

Acoustics 74.7 (2013), pp. 921–925. ISSN: 0003682X. DOI: [10.1016/j.apacoust.2013.01.010](#) (cit. on p. 13).

[Hof14]   Hoffman, Matthew D.; Gelman, Andrew et al.: "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo". In: *J. Mach. Learn. Res.* 15.1 (2014), pp. 1593–1623 (cit. on p. 136).

[Hub62]   Hubel, David H. and Wiesel, Torsten N.: "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1 (1962), p. 106 (cit. on p. 33).

[Hut19]   Hutter, Frank; Kotthoff, Lars and Vanschoren, Joaquin, eds.: Automated Machine Learning: Methods, Systems, Challenges. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-05318-5 (cit. on p. 85).

[Ian16]   Ian Goodfellow; Yoshua Bengio and Aaron Courville: Deep Learning. MIT press, 2016. URL: [http : / / www . deeplearningbook.org](#) (visited on 01/05/2024) (cit. on pp. 33, 36).

[Int13]   International Electrotechnical Commission: Electroacoustics - Sound level meters - Part 1: Specifications. 2013 (cit. on p. 20).

[Int19]   International Organization for Standardization: Tyres - Coast-by methods for measurement of tyre-to-road sound emission. 2019 (cit. on p. 4).

[Int21]   International Organization for Standardization: Acoustics - Measurement of the influence of road surfaces on traffic noise - Part 3: Reference tyres. 2021. (Visited on 10/11/2023) (cit. on p. 8).

[Iva21]   Ivanovs, Maksims; Kadikis, Roberts and Ozols, Kaspars: "Perturbation-based methods for explaining deep neural networks: A survey". In: *Pattern Recognition Letters* 150 (2021), pp. 228–234. ISSN: 01678655. DOI: [10.1016/j.patrec.2021.06.030](#) (cit. on p. 42).

[Iwa96]   Iwao, Keijiro and Yamazaki, Ichiro: "A study on the mechanism of tire/road noise". In: *JSAE review* 17.2 (1996), pp. 139–144 (cit. on p. 17).

[Jab12]   Jabben, Jan; Verheijen, Edwin and Potma, Charlos: "Noise reduction by electric vehicles in the Netherlands". In: *Institute of Noise Control Engineering 2012 – Inter-Noise and Noise-Con Congress.* 2012, pp. 6958–6965 (cit. on pp. 1, 3).

[Jif17]   Jifeng Dai; Haozhi Qi; Yuwen Xiong; Yi Li; Guodong Zhang; Han Hu and Yichen Wei: Deformable Convolutional Networks. 2017. url: https://arxiv.org/pdf/1703.06211.pdf (visited on 01/25/2024) (cit. on p. 35).

[Kap21]   Kapishnikov, Andrei; Venugopalan, Subhashini; Avci, Besim; Wedin, Ben; Terry, Michael and Bolukbasi, Tolga: "Guided Integrated Gradients: an Adaptive Path Method for Removing Noise". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE, 2021, pp. 5048–5056. isbn: 978-1-6654-4509-2. doi: 10.1109/CVPR46437.2021.00501 (cit. on p. 45).

[Ket17]   Ketkar, Nikhil: Deep Learning with Python. Berkeley, CA: Apress, 2017. isbn: 978-1-4842-2765-7. doi: 10.1007/978-1-4842-2766-4 (cit. on p. 36).

[Kim06]   Kim, Sungtae; Jeong, Wontae; Park, Yonghwan and Lee, Soogab: "Prediction method for tire air-pumping noise using a hybrid technique". In: *The Journal of the Acoustical Society of America* 119.6 (2006), pp. 3799–3812. issn: 0001-4966. doi: 10.1121/1.2200140 (cit. on p. 21).

[Kin08]   Kingsford, Carl and Salzberg, Steven L.: "What are decision trees?" In: *Nature biotechnology* 26.9 (2008), pp. 1011–1013. doi: 10.1038/nbt0908-1011 (cit. on p. 53).

[Kle08]   KLEIN, Philippe; BECKENBAUER, Thomas; HAMET, Jean-Francois
           and KROPP, Wolfgang: "Tyre-road noise prediction: A compari-
           son between the SPERoN and HyRoNE models-Part 2". In: *Jour-
           nal of the Acoustical Society of America* 123.5 (2008), p. 3389 (cit.
           on p. 22).

[Kok20]   KOKHLIKYAN, Narine et al.: Captum: A unified and generic
           model interpretability library for PyTorch. 2020. URL: http :
           //arxiv.org/pdf/2009.07896v1 (visited on 01/25/2024) (cit. on
           p. 42).

[Kro08]   KROGH, Anders: "What are artificial neural networks?" In: *Na-
           ture biotechnology* 26.2 (2008), pp. 195–197 (cit. on p. 29).

[Kro91]   KROGH, Anders and HERTZ, John: "A simple weight decay can
           improve generalization". In: *Advances in neural information pro-
           cessing systems* 4 (1991) (cit. on p. 31).

[Kum11]   KUMAR, A.; TANDON, A.; PAUL, S.; SINGLA, A.; KUMAR, S.; VIJAY,
           P. and BHANGALE, U. D.: "Influence of tyre's dimensional charac-
           teristics on tyre-pavement noise emission". In: *Physical Review
           and Research International* 1.4 (2011), pp. 124–137 (cit. on pp. 5–
           7).

[Lee08]   LEE, Tae Keun and KIM, Byoung Sam: "Vibration analysis of
           automobile tire due to bump impact". In: *Applied Acoustics* 69.6
           (2008), pp. 473–478. ISSN: 0003682X. DOI: 10.1016/j.apacoust.
           2007.12.003 (cit. on p. 9).

[Lee21]   LEE, Sang-Kwon; LEE, Hwajin; BACK, Jiseon; AN, Kanghyun;
           YOON, Youngsam; YUM, Kiho; KIM, Sungdae and HWANG, Sung-
           Uk: "Prediction of tire pattern noise in early design stage based
           on convolutional neural network". In: *Applied Acoustics* 172
           (2021), p. 107617. ISSN: 0003682X. DOI: 10.1016/j.apacoust.
           2020.107617. (Visited on 12/09/2020) (cit. on p. 23).

[Leu22]   LEUPOLZ, Michael and GAUTERIN, Frank: "Vehicle Impact on
           Tire Road Noise and Validation of an Algorithm to Virtually
           Change Tires". In: *Applied Sciences* 12.17 (2022), p. 8810. DOI:
           10.3390/app12178810 (cit. on pp. 14, 20, 50, 66, 68, 69, 73, 75, 83).

[Leu23]   Leupolz, Michael; Radler, Timotheus and Gauterin, Frank: "Impact of tessellation surface model generation process on acoustical tire features". In: *Measurement* 219 (2023), p. 113225. ISSN: 02632241. DOI: 10.1016/j.measurement.2023.113225 (cit. on pp. 55, 57, 61, 131).

[Li16]   Li, Tan; Burdisso, Ricardo and Sandu, Corina: "The effects of tread patterns on tire pavement interaction noise". In: *Institute of Noise Control Engineering 2016 – Inter-Noise and Noise-Con Congress*. Vol. 253. 2016, pp. 208–219 (cit. on pp. 12, 61).

[Li17]   Li, Tan; Burdisso, Ricardo and Sandu, Corina: "An Artificial Neural Network Model to Predict Tread Pattern-Related Tire Noise". In: *SAE Technical Paper Series*. SAE Technical Paper Series. SAE International400 Commonwealth Drive, Warrendale, PA, United States, 2017. DOI: 10.4271/2017-01-1904 (cit. on pp. 22, 24, 62).

[Li18a]   Li, Tan: "Influencing Parameters on Tire–Pavement Interaction Noise: Review, Experiments and Design Considerations". In: *Designs* 2.4 (2018), p. 38. DOI: 10.3390/designs2040038 (cit. on p. 3).

[Li18b]   Li, Tan; Burdisso, Ricardo and Sandu, Corina: "Literature review of models on tire-pavement interaction noise". In: *Journal of Sound and Vibration* 420 (2018), pp. 357–445. ISSN: 0022460X. DOI: 10.1016/j.jsv.2018.01.026 (cit. on p. 25).

[Li21]   Li, Zewen; Liu, Fan; Yang, Wenjie; Peng, Shouheng and Zhou, Jun: "A survey of convolutional neural networks: analysis, applications, and prospects". In: *IEEE transactions on neural networks and learning systems* (2021) (cit. on p. 36).

[Lil08]   Liljegren, Karin: "Visual and acoustic tyre tread design". Master's Thesis. Göteburg, Sweden: Chalmers University of Technology, 2008 (cit. on p. 12).

[Lin06]   Linder, Wilfried: Digital photogrammetry: A practical course. 2nd ed. Berlin and New York: Springer, 2006. ISBN: 3-540-29152-0. (Visited on 04/18/2023) (cit. on p. 55).

[Lip01]   Lipovetsky, Stan and Conklin, Michael: "Analysis of regression in game theory approach". In: *Applied Stochastic Models in Business and Industry* 17.4 (2001), pp. 319–330. ISSN: 1524-1904. DOI: 10.1002/asmb.446 (cit. on p. 46).

[Lun17]   Lundberg, Scott M. and Lee, Su-In: "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems.* Ed. by I. Guyon; U. Von Luxburg; S. Bengio; H. Wallach; R. Fergus; S. Vishwanathan and R. Garnett. Vol. 30. Curran Associates, Inc, 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (cit. on pp. 46, 100).

[Lyn07]   Lynch, Scott M.: Introduction to Applied Bayesian Statistics and Estimation for Social Scientists. New York, NY: Springer New York, 2007. ISBN: 978-0-387-71264-2. DOI: 10.1007/978-0-387-71265-9 (cit. on p. 38).

[Mar85]   Marshek, Kurt M.; Hudson, W. Ronald; Connell, Richard B.; Chen, Hsien H. and Saraf, Chhote L.: "Experimental Investigation of Truck Tire Inflation Pressure on Pavement-Tire Contact Area and Pressure Distribution: Research Report Number 386-1". In: (1985). URL: https://api.semanticscholar.org/CorpusID:106867698 (visited on 10/10/2023) (cit. on p. 16).

[Med01]   Medsker, Larry R. and Jain, L. C.: "Recurrent neural networks". In: *Design and Applications* 5.64-67 (2001), p. 2 (cit. on p. 30).

[Mig20]   Miglani, Vivek; Kokhlikyan, Narine; Alsallakh, Bilal; Martin, Miguel and Reblitz-Richardson, Orion: Investigating Saturation Effects in Integrated Gradients. 2020. URL: http://arxiv.org/pdf/2010.12697v1 (visited on 01/25/2024) (cit. on p. 45).

[Mik01]   Mikhail, Edward M.; Bethel, James S. and McGlone, J. Chris: Introduction to modern photogrammetry. John Wiley & Sons, 2001. ISBN: 978-0-471-30924-6. (Visited on 04/18/2023) (cit. on p. 55).

[Moh18]   MOHRI, Mehryar; ROSTAMIZADEH, Afshin and TALWALKAR,
          Ameet: Foundations of machine learning. MIT press, 2018. ISBN:
          9780262039406 (cit. on pp. 27, 30, 31, 40, 41).

[Moh22]   MOHAMMADI, Somaye; OHADI, Abdolreza and IRANNEJAD-
          PARIZI, Mostafa: "A comprehensive study on statistical predic-
          tion and reduction of tire/road noise". In: *Journal of Vibration
          and Control* 28.19-20 (2022), pp. 2487–2501. ISSN: 1077-5463. DOI:
          10.1177/10775463211013184. (Visited on 02/09/2023) (cit. on
          pp. 24, 62).

[Mol22]   MOLNAR, Christoph: Interpretable machine learning: A
          guide for making black box models explainable. Second
          edition. Munich, Germany: Christoph Molnar, 2022. URL:
          https://christophm.github.io/interpretable-ml-book/ (cit. on
          pp. 46, 47).

[Mon12]   MONTGOMERY, Douglas C.; PECK, Elizabeth A. and VINING,
          G. Geoffrey: Introduction to linear regression analysis. 5. ed.
          Wiley series in probability and statistics. Hoboken, NJ: Wiley,
          2012. ISBN: 978-0-470-54281-1 (cit. on pp. 27, 28).

[Mun14]   MUNIANDY, Ratnasamy; MOAZAMI, Danial; HAMID, Hussain and
          HASSIM, Salihudin: "Characterization of Effective Tire Contact
          Area for Various Tread Patterns". In: *Instrumentation Science &
          Technology* 42.1 (2014), pp. 15–26. ISSN: 1073-9149. DOI: 10.1080/
          10739149.2013.832291 (cit. on p. 16).

[Nil76]   NILSSON, Nils-Åke: On generating mechanisms for external tire
          noise. 1976 (cit. on pp. 13, 75).

[Nor04]   NORDSTRÖM, Olle: Friction on ice for new and used winter tyres.
          Summary report An investigation concerning the influence of
          age, tread depth, tread rubber hardness, stud protrusion and
          stud force. Statens väg-och transportforskningsinstitut., VTI
          meddelande 966, 2004 (cit. on p. 8).

[Obr07]   O'brien, Robert M.: "A Caution Regarding Rules of Thumb for Variance Inflation Factors". In: *Quality & Quantity* 41.5 (2007), pp. 673–690. issn: 0033-5177. doi: [10.1007/s11135-006-9018-6](https://doi.org/10.1007/s11135-006-9018-6) (cit. on p. 27).

[Oor17]   Oorath, Rahul; Saraswat, Abhishek; Gupta, Shivashish Kumar; Tiwari, Nachiketa; Goyal, Sharad and Patel, Chirag: "Influence of tread design parameters on air pumping noise in automotive tires". In: *24th International Congress on Sound and Vibration* (2017) (cit. on p. 23).

[OSh15]   O'Shea, Keiron and Nash, Ryan: An introduction to convolutional neural networks. 2015. url: https://arxiv.org/pdf/1511.08458.pdf (visited on 01/25/2024) (cit. on pp. 33, 35, 36).

[Pas19]   Paszke, Adam; Gross, Sam; Massa, Francisco; Lerer, Adam; Bradbury, James; Chanan, Gregory; Killeen, Trevor; Lin, Zeming; Gimelshein, Natalia; Antiga, Luca et al.: "Pytorch: An imperative style, high-performance deep learning library". In: *Advances in neural information processing systems* 32 (2019), pp. 8024–8035. url: https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf (visited on 01/25/2024) (cit. on p. 86).

[Ped11]   Pedregosa, F. et al.: "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830. url: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html (cit. on pp. 54, 79).

[Pin20]   Pinay, Julien; Saito, Yoshinori; Mignot, Christian and Gauterin, Frank: "Understanding the contribution of groove resonance to tire-road noise on different surfaces under various operating conditions". In: *Acta Acustica* 4.2 (2020), p. 6. doi: [10.1051/aacus/2020004](https://doi.org/10.1051/aacus/2020004) (cit. on p. 12).

[Pin23]   Pinay, Julien: "Experimental investigation of relevant road surface descriptors for tire-road noise measurements on low-absorbing road surfaces". PhD thesis. 2023. doi: [10.5445/KSP/1000163480](https://doi.org/10.5445/KSP/1000163480) (cit. on p. 3).

[Pol17]   Polasik, Jakub; Waluś, Konrad J. and Warguła, Łukasz: "Experimental Studies of the Size Contact Area of a Summer Tire as a Function of Pressure and the Load". In: *Procedia Engineering* 177 (2017), pp. 347–351. issn: 18777058. doi: [10.1016/j.proeng.2017.02.203](https://doi.org/10.1016/j.proeng.2017.02.203) (cit. on p. 15).

[Pot17]   Potdar, Kedar; Pardawala, Taher S. and Pai, Chinmay D.: "A comparative study of categorical variable encoding techniques for neural network classifiers". In: *International journal of computer applications* 175.4 (2017), pp. 7–9 (cit. on p. 54).

[Pou19]   Pouyanfar, Samira; Sadiq, Saad; Yan, Yilin; Tian, Haiman; Tao, Yudong; Reyes, Maria Presa; Shyu, Mei-Ling; Chen, Shu-Ching and Iyengar, S. S.: "A Survey on Deep Learning". In: *ACM Computing Surveys* 51.5 (2019), pp. 1–36. issn: 0360-0300. doi: [10.1145/3234150](https://doi.org/10.1145/3234150) (cit. on p. 36).

[Put13]   Putner, Jakob; Lohrmann, Martin and Fastl, Hugo: "Contribution analysis of vehicle exterior noise with operational transfer path analysis". In: *Proceedings of Meetings on Acoustics ICA2013*. Vol. 19. 2013, p. 040035 (cit. on pp. 1, 3).

[Rad22]   Radler, Timotheus: "Development of a tool for generating a surface model from a series of images of passenger car tires". Master's Thesis. Karlsruhe, Germany: Karlsruhe Institue of Technology, 2022 (cit. on p. 55).

[Rap23]   Rapino, Luca; Liu, Ling; Dinosio, Arianna; Ripamonti, Francesco; Corradi, Roberto and Baro, Simone: "Processing of tyre data for rolling noise prediction through a statistical modelling approach". In: *Mechanical Systems and Signal Processing* 188 (2023), p. 110042. doi: [10.1016/j.ymssp.2022.110042](https://doi.org/10.1016/j.ymssp.2022.110042) (cit. on p. 24).

[Rie19]   Riehm, Patrick; Unrau, Hans-Joachim; Gauterin, Frank; Torbrügge, Stefan and Wies, Burkhard: "3D brush model to predict longitudinal tyre characteristics". In: *Vehicle System Dynamics* 57.1 (2019), pp. 17–43. issn: 0042-3114. doi: [10.1080/00423114.2018.1447135](https://doi.org/10.1080/00423114.2018.1447135) (cit. on p. 15).

[Rok05]   Rokach, Lior and Maimon, Oded: "Decision Trees". In: *Data Mining and Knowledge Discovery Handbook.* Ed. by Maimon, Oded and Rokach, Lior. Boston, MA: Springer US, 2005, pp. 165–192. isbn: 978-0-387-25465-4. doi: 10.1007/0-387-25465-X_9 (cit. on p. 53).

[Roz22]   Rozemberczki, Benedek; Watson, Lauren; Bayer, Péter; Yang, Hao-Tsung; Kiss, Olivér; Nilsson, Sebastian and Sarkar, Rik: The Shapley Value in Machine Learning. 2022. url: http://arxiv.org/pdf/2202.05594v2 (visited on 01/25/2024) (cit. on p. 47).

[Sae08]   Saemann, Ernst-Ulrich: "Contribution of the tyre to further lowering tyre-road noise". In: *Journal of the Acoustical Society of America* 123.5 (2008), p. 3867 (cit. on pp. 8, 11).

[Sal15]   Salimans, Tim; Kingma, Diederik and Welling, Max: "Markov chain monte carlo and variational inference: Bridging the gap". In: *International conference on machine learning.* 2015, pp. 1218–1226 (cit. on p. 38).

[San02]   Sandberg, Ulf and Ejsmont, Jerzy A.: Tyre/road noise: Reference book. Kisa, Sweden: Informex, 2002. isbn: 91-631-2610-9 (cit. on pp. 3–14, 17, 66, 75, 103, 125, 138).

[San07]   Sandberg, Ulf and Ejsmont, J.: "Influence of tyre tread rubber hardness on tyre/road noise emission". In: *Inter-Noise, Istanbul, Turkey* (2007) (cit. on p. 8).

[San17]   Sandberg, Ulf: "Calibrating the ISO 10844 test surfaces used for vehicle and tyre noise testing". In: *46th International Congress and Exposition on Noise Control Engineering: Taming Noise and Moving Quiet, INTER-NOISE 2017, 27 August 2017 through 30 August 2017.* 2017 (cit. on p. 2).

[San21]   Sandberg, Ulf; Berge, Truls and Schlatter, Felix, Bühlmann, Erik: Improving reproducibility of tyre/road noise measurements on ISO test tracks: CEDR Transnational Road Research Programme Call 2018: Noise and Nuisance. 2021 (cit. on p. 2).

[San80]   SANDBERG, Ulf and DESCORNET, Guy: Road surface influence on tire/road noise. 1980 (cit. on p. 3).

[Sch15]   SCHMIDHUBER, Juergen: "Deep Learning in Neural Networks: An Overview". In: *Neural Networks* 61.3 (2015), pp. 85–117. ISSN: 08936080. DOI: 10.1016/j.neunet.2014.09.003. URL: http://arxiv.org/pdf/1404.7828v4 (cit. on p. 36).

[Sch17]   SCHNIEDERS, Lars: "Leisere Antriebsreifen für den Regionalverkehr". In: *Leiser Straßenverkehr 3*. 2017, pp. 147–176 (cit. on p. 11).

[Sha17]   SHARMA, Sagar; SHARMA, Simone and ATHAIYA, Anidhya: "Activation functions in neural networks". In: *Towards Data Sci* 6.12 (2017), pp. 310–316 (cit. on p. 31).

[Sha53]   SHAPLEY, L. S.: "A Value for n-Person Games". In: *Contributions to the Theory of Games (AM-28), Volume II*. Ed. by KUHN, Harold William and TUCKER, Albert William. Princeton University Press, 1953, pp. 307–318. ISBN: 9781400881970. DOI: 10.1515/9781400881970-018 (cit. on pp. 45–47).

[Sha96]   SHARMA, Ajay Kumar and PANDEY, K. P.: "A review on contact area measurement of pneumatic tyre on rigid and deformable surfaces". In: *Journal of Terramechanics* 33.5 (1996), pp. 253–264. ISSN: 0022-4898. DOI: 10.1016/S0022-4898(97)00008-6. URL: https://www.sciencedirect.com/science/article/pii/S0022489897000086 (cit. on p. 15).

[Smi15]   SMITH, Leslie N.: Cyclical Learning Rates for Training Neural Networks. 2015. URL: http://arxiv.org/pdf/1506.01186v6 (visited on 01/25/2024) (cit. on p. 90).

[Smi18]   SMITH, Leslie N.: A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. 2018. URL: https://arxiv.org/pdf/1803.09820.pdf (visited on 01/25/2024) (cit. on p. 31).

[Son17]   SONODA, Sho and MURATA, Noboru: "Neural network with un-
          bounded activation functions is universal approximator". In: *Ap-
          plied and Computational Harmonic Analysis* 43.2 (2017), pp. 233–
          268. ISSN: 10635203. DOI: 10.1016/j.acha.2015.12.005 (cit. on
          p. 30).

[Štr14]   ŠTRUMBELJ, Erik and KONONENKO, Igor: "Explaining prediction
          models and individual predictions with feature contributions".
          In: *Knowledge and Information Systems* 41.3 (2014), pp. 647–665.
          ISSN: 0219-1377. DOI: 10.1007/s10115-013-0679-x (cit. on pp. 46,
          47).

[Stu20]   STURMFELS, Pascal; LUNDBERG, Scott and LEE, Su-In: "Visualiz-
          ing the Impact of Feature Attribution Baselines". In: *Distill* 5.1
          (2020). ISSN: 2476-0757. DOI: 10.23915/distill.00022 (cit. on p. 44).

[Sun17]   SUNDARARAJAN, Mukund; TALY, Ankur and YAN, Qiqi: Ax-
          iomatic Attribution for Deep Networks. 2017. URL: http :
          //arxiv.org/pdf/1703.01365v2 (visited on 01/25/2024) (cit. on
          pp. 43, 44).

[Sun18]   SUNDARARAJAN, Mukund and TALY, Ankur: A Note about: Local
          Explanation Methods for Deep Neural Networks lack Sensitivity
          to Parameter Values. 2018. URL: http://arxiv.org/pdf/1806.
          04205v1 (visited on 01/25/2024) (cit. on p. 44).

[The11]   THEAKSTON, Frank, ed.: Burden of disease from environmental
          noise: Quantification of healthy life years lost in Europe. Copen-
          hagen: World Health Organization Regional Office for Europe,
          2011. ISBN: 9789289002295 (cit. on p. 1).

[Tib96]   TIBSHIRANI, Robert: "Regression shrinkage and selection via
          the lasso". In: *Journal of the Royal Statistical Society Series B:
          Statistical Methodology* 58.1 (1996), pp. 267–288 (cit. on p. 29).

[Ulf04]   ULF SANDBERG: "A Closer Look at the Tread Groove Resonance
          in Tyre/Road Noise". In: 2004. URL: https://api.semanticscholar.
          org/CorpusID:207969680 (cit. on p. 16).

[UNE18]   UN/ECE Regulation No 51: Uniform provisions concerning the approval of motor vehicles having at least four wheels with regard to their sound emissions [2018/ 798]. 2018. (Visited on 04/08/2022) (cit. on pp. 2, 68).

[UNE22]   UN/ECE Regulation No 117: Uniform provisions concerning the approval of tyres with regard to rolling sound emissions and to adhesion on wet surfaces and/or to rolling resistance. 2022. (Visited on 04/08/2022) (cit. on pp. 2, 19, 20, 68, 94).

[UNE23a]   UNECE: (IWG MU) Proposal for UN Regulation No. R117 C2 noise temperature correction update. 2023. URL: https://unece. org/transport/documents/2023/07/informal-documents/iwg-mu-proposal-un-regulation-no-r117-c2-noise (visited on 10/12/2023) (cit. on p. 20).

[UNE23b]   UNECE: Working Party on Noise and Tyres (78th session). 2023. URL: https://unece.org/transport/events/wp29grbp-working-party-noise-and-tyres-78th-session (visited on 12/23/2023) (cit. on p. 20).

[van05]   van Keulen, W. and Duskov, M.: Inventory study of basic Inventory study of basic knowledge on tyre/road noise: Report, Poject report DWW-2005-022. Ed. by IPG. 2005 (cit. on p. 1).

[van14]   van der Walt, Stéfan; Schönberger, Johannes L.; Nunez-Iglesias, Juan; Boulogne, François; Warner, Joshua D.; Yager, Neil; Gouillart, Emmanuelle; Yu, Tony and the scikit-image contributors: "scikit-image: image processing in Python". In: PeerJ 2 (2014), e453. ISSN: 2167-8359. DOI: 10.7717/peerj.453 (cit. on p. 129).

[van21]   van de Schoot, Rens et al.: "Bayesian statistics and modelling". In: Nature Reviews Methods Primers 1.1 (2021). DOI: 10.1038/s43586-020-00001-2 (cit. on pp. 37, 38).

[Vee23]   Veenman, Myrthe; Stefan, Angelika M. and Haaf, Julia M.: "Bayesian hierarchical modeling: an introduction and reassessment". In: Behavior research methods (2023). DOI: 10.3758/s13428-023-02204-3 (cit. on p. 38).

[Was04]   WASSERMAN, Larry: All of Statistics. New York, NY: Springer New York, 2004. ISBN: 978-1-4419-2322-6. DOI: 10.1007/978-0-387-21736-9 (cit. on p. 37).

[Weh18]   WEHR, Reinhard; FUCHS, Andreas and AICHINGER, Claus: "A combined approach for correcting tyre hardness and temperature influence on tyre/road noise". In: *Applied Acoustics* 134 (2018), pp. 110–118. ISSN: 0003682X. DOI: 10.1016/j.apacoust.2018.01.004 (cit. on p. 9).

[Wei09]   WEINBERGER, Kilian; DASGUPTA, Anirban; LANGFORD, John; SMOLA, Alex and ATTENBERG, Josh: "Feature Hashing for Large Scale Multitask Learning". In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 1113–1120. ISBN: 9781605585161. DOI: 10.1145/1553374.1553516 (cit. on p. 54).

[Wil03]   WILSON, D. Randall and MARTINEZ, Tony R.: "The general inefficiency of batch training for gradient descent learning". In: *Neural Networks* 16.10 (2003), pp. 1429–1451. ISSN: 08936080. DOI: 10.1016/S0893-6080(03)00138-2 (cit. on p. 33).

[Xu08]   XU, Huan; CARAMANIS, Constantine and MANNOR, Shie: "Robust regression and lasso". In: *Advances in neural information processing systems* 21 (2008) (cit. on p. 29).

[Yan13]   YANG, Jia Sheng; FWA, Tien Fang; ONG, Ghim Ping and CHEW, Chye Heng: "Finite-Element Analysis of Effect of Wide-Base Tire on Tire-Pavement Noise". In: *Advanced Materials Research* 723 (2013), pp. 105–112. DOI: 10.4028/www.scientific.net/AMR.723.105 (cit. on pp. 6, 125).

[Zel18]   ZELLER, Peter, ed.: Handbuch Fahrzeugakustik: Grundlagen, Auslegung, Berechnung, Versuch. 3. ATZ/MTZ-Fachbuch. Wiesbaden and Heidelberg: Springer Vieweg, 2018. ISBN: 978-3-658-18519-0. DOI: 10.1007/978-3-658-18520-6 (cit. on pp. 1, 3).

[Zho21]   Zhou, Zhi-Hua: Machine Learning. Singapore: Springer Singapore, 2021. isbn: 978-981-15-1966-6. doi: 10.1007/978-981-15-1967-3 (cit. on p. 29).

Supervised student theses

These theses were not included in this dissertation.

[1]   EBERLE, Lukas: "Objektivierung & Klasifizierung von BSR-Störgeräuschen im PKW-Interieur / Objectification & Classification of BSR Noise in Passenger Car Interiors". Master's Thesis. Karlsruhe Institute of Technology, 2021.

[2]   AL ARABI, Rami: "Sound Intensity Analysis of Passenger Vehicle Tires on Different Surfaces / Analyse der Schallintensität von PKW-Reifen auf verschiedenen Untergründen". Master's Thesis. Karlsruhe Institute of Technology, 2022.

[3]   RADLER, Timotheus: "Erstellung eines Tools zur Generierung eines Flächenmodells aus einer Bilderserie von PKW-Reifen / Development of a tool for generating a surface model from a series of images of passenger car tires". Master's Thesis. Karlsruhe Institute of Technology, 2022.

[4]   LITTERST, Nico: "Entwurf und Umsetzung einere Regelung zur optimierten Rotorpositionierung für elektrische Antriebe / Design and implementation of a control system for optimised rotor positioning for electric drives". Masters's Thesis. Karlsruhe Institute of Technology, 2023.

[5]   SIEHL, Pascal: "Komfortvergleich zwischen Ein- und Zwei-Ventil-Dämpfern in semiaktiven Fahrwerken / Driving Comfort Comparison of One-Valve and Two-Valve Shock Absorbers in Semi-Active Suspensions". Master's Thesis. Karlsruhe Institute of Technology, 2023.

List of Figures

List of Tables

A   Additional Data

Table A.1: Overview on all used input with min and mac values and dimensions

Feature	Minimum	Maximum	Dimension
Tire nominal width	205	285	mm
Tire aspect ratio	30	65	$100 \times \%$
Rim diameter	16	23	in
Outer diameter	641.60	815.30	mm
Tire load index	615	1180	kg
Tire speed index	190	300	$\mathrm{km\,h^{-1}}$
Shore hardness block	50.975	73.95	°Sh
Tire temperature block	18.25	25.75	°C
Shore hardness shoulder	56.675	74.275	°Sh
Tire temperature shoulder	17.5	25.75	°C
Tire profile depth	5.525	8.475	mm
Tread block-to-groove ratio	0.48	0.70	%
Avg. circumferential groove width	7.83	15.35	mm
Tire spectrum	-	-	
Vehicle speed	32.68	101.09	$\mathrm{km\,h^{-1}}$
Measurement vehicle	-	-	
Driving direction	-	-	
Tire manufacturer	-	-	
Tire fragment images	-	-	
Tire fragment images (Fuji)	-	-	
Air temperature	13.40	31.47	°C
Road temperature	15.82	49.37	°C
Wind speed	0.0	17.94	$\mathrm{km\,h^{-1}}$
Wind direction	1.60	359.93	°
Air pressure	929.45	944.38	kPa
Humidity	25.04	94.21	%
Distance	-10	10	m

Table A.2: Overview of data used in different models

Feature	TRN	Linear Reg.	Lasso Reg.	ANN	CNN	CM	CM$^{\text{Distance}}$	CM$^{\text{Distance}}_{\text{Avg}}$	CM$^{\text{Distance}}_{\text{Avg, Fuji}}$	CM$^{\text{PSD}}$	Bayes$^{\text{PP}}_{\text{simple}}$	Bayes$^{\text{PP}}_{\text{complex}}$
Tire nominal width		×	×	×		×	×	×	×	×		
Tire aspect ratio		×	×	×		×	×	×	×	×		
Rim diameter		×	×	×		×	×	×	×	×		
Tire load index		×	×	×		×	×	×	×	×		
Tire speed index		×	×	×		×	×	×	×	×		
Shore hardness block		×	×	×		×	×	×	×	×		
Tire temperature block		×	×	×		×	×	×	×	×		
Shore hardness shoulder		×	×	×		×	×	×	×	×		
Tire temperature shoulder		×	×	×		×	×	×	×	×		
Tire profile depth												
Tread block-to-groove ratio											×	×
Average circ. groove width											×	×
Tire spectrum										×		
Vehicle speed	×	×	×	×	×	×	×	×	×	×		
Measurement vehicle				×		×	×	×	×	×	×	×
Tire manufacturer				×		×	×	×	×	×		
Tire fragment images					×							
Tire fragment images (fuji)									×			
Air temperature		×	×			×	×	×	×	×	×	×
Road temperature		×	×			×	×	×	×	×		
Wind speed		×	×			×	×	×	×	×		
Wind direction		×	×			×	×	×	×	×		
Air pressure		×	×			×	×	×	×	×		
Humidity		×	×			×	×	×	×	×		
Distance						×	×	×	×	×		×

Table A.3: Combined CNN and ANN Leave one Out performance metric for full test track prediction. All values expect R^2 values in dB(A).

Model	Mean$_{\text{RMSE}}$	STD$_{\text{RMSE}}$	Max$_{\text{RMSE}}$	Min$_{\text{RMSE}}$	Mean$_{\text{RMSE, train}}$
TRN coefficients	1.56	0.78	3.60	0.6	$R^2 = 0.89$
Linear Regression	1.54	0.65	3.04	0.68	$R^2 = 0.90$
Lasso Regression	1.45	0.55	2.95	0.70	$R^2 = 0.91$
ANN	1.0	0.31	1.84	0.64	0.79
ANN$_{\text{Test optimized}}$	0.86	0.26	1.57	0.51	1.30
CNN$^{\sim 80\,\text{km h}^{-1}}$	0.73	0.33	2.10	0.28	0.74
CNN$^{\sim 80\,\text{km h}^{-1}}_{\text{Test optimized}}$	0.53	0.17	0.86	0.18	1.67
CM	0.95	0.28	1.59	0.51	0.63
CM$_{\text{Test optimized}}$	0.82	0.26	1.50	0.49	1.04
CM$^{\text{Distance}}$	1.08	0.23	1.55	0.66	0.62
CM$^{\text{Distance}}_{\text{Test optimized}}$	0.98	0.21	1.55	0.60	1.00
CM$^{\text{Distance}}_{\text{tire aspect ratio}}$	1.14	0.22	1.57	0.65	0.62
CM$^{\text{Distance}}_{\text{rim diameter}}$	1.17	0.24	1.64	0.79	0.60
CM$^{\text{Distance}}_{\text{outer diameter}}$	1.20	0.24	1.69	0.74	0.63
CM$^{\text{Distance}}_{\text{Avg}}$	0.64	0.16	0.93	0.34	0.39
CM$^{\text{Distance}}_{\text{Avg ,Test optimized}}$	0.60	0.16	0.92	0.28	0.50
CM$^{\text{Distance}}_{\text{Avg, Fuji}}$	0.58	0.15	0.98	0.32	0.37
CM$^{\text{Distance}}_{\text{Avg Fuji ,Test optimized}}$	0.51	0.13	0.98	0.32	0.59
CM$^{\text{PSD}}_{1/3\,\text{octave}}$	24.83	3.71	33.19	18.73	13.72
CM$^{\text{PSD}}_{\text{Full}}$	23.86	3.72	32.29	18.18	11.19
CM$^{\text{PSD}}_{\text{Embedded}}$	23.31	3.60	30.22	17.89	13.75
Bayes$^{\text{PP'}}_{\text{simple}}$	0.97	0.60	2.79	0.33	1.02
Bayes$^{\text{PP'}}_{\text{complex}}$	1.16	0.64	2.88	0.28	0.28

Table A.4: Combined CNN and ANN Leave one Out performance metric for full test track prediction. All values expect R^2 values in dB(A). The RMSE values for this table are averaged over the number of output variables of the models.

Model	$\text{Mean}_{\text{RMSE}}$	STD_{RMSE}	Max_{RMSE}	Min_{RMSE}	$\text{Mean}_{\text{RMSE, train}}$
TRN coefficients	1.11	0.55	2.54	0.42	$R^2 = 0.89$
Linear Regression	1.09	0.46	2.15	0.48	$R^2 = 0.90$
Lasso Regression	1.03	0.39	2.08	0.50	$R^2 = 0.91$
ANN	0.71	0.22	1.30	0.45	0.56
$\text{ANN}_{\text{Test optimized}}$	0.61	0.18	1.11	0.37	0.92
$\text{CNN}^{\sim 80\,\text{km h}^{-1}}$	0.52	0.23	1.48	0.20	0.52
$\text{CNN}^{\sim 80\,\text{km h}^{-1}}_{\text{Test optimized}}$	0.38	0.11	0.62	0.12	1.18
CM	0.67	0.20	1.12	0.36	0.44
$\text{CM}_{\text{Test optimized}}$	0.58	0.18	1.06	0.35	0.73
$\text{CM}^{\text{Distance}}$	0.76	0.16	1.09	0.47	0.44
$\text{CM}^{\text{Distance}}_{\text{Test optimized}}$	0.70	0.15	1.09	0.42	0.70
$\text{CM}^{\text{Distance}}_{\text{tire aspect ratio}}$	0.80	0.15	1.11	0.46	0.44
$\text{CM}^{\text{Distance}}_{\text{rim diameter}}$	0.82	0.17	1.16	0.56	0.42
$\text{CM}^{\text{Distance}}_{\text{outer diameter}}$	0.85	0.17	1.19	0.52	0.45
$\text{CM}^{\text{Distance}}_{\text{Avg}}$	0.64	0.16	0.93	0.34	0.39
$\text{CM}^{\text{Distance}}_{\text{Avg ,Test optimized}}$	0.60	0.16	0.92	0.28	0.50
$\text{CM}^{\text{Distance}}_{\text{Avg, Fuji}}$	0.58	0.15	0.98	0.32	0.37
$\text{CM}^{\text{Distance}}_{\text{Avg Fuji ,Test optimized}}$	0.51	0.13	0.98	0.32	0.59
$\text{CM}^{\text{PSD}}_{\text{1/3 octave}}$	3.15	0.47	4.22	2.38	1.74
$\text{CM}^{\text{PSD}}_{\text{Full}}$	3.03	0.47	4.10	2.31	1.42
$\text{CM}^{\text{PSD}}_{\text{Embedded}}$	2.96	0.46	3.83	2.27	1.75
$\text{Bayes}^{\text{PP'}}_{\text{simple}}$	0.97	0.60	2.79	0.33	1.02
$\text{Bayes}^{\text{PP'}}_{\text{complex}}$	1.16	0.64	2.88	0.28	0.28

Table A.5: List of all tires included in measurements with vehicle with which measurements are performed.

Manufacturer	Profile	Dimensions	a	b	R^2
Dunlop	Sport BlueResponse	225/55R19 95V	10.63	32.13	0.9928
Goodyear	Wrangler HP	215/60R16 95H	6.90	34.35	0.9944
Michelin	Energy Saver MO	205/65R16 95V	7.94	33.56	0.9928
Bridgestone	Alenza 001 MO	235/45R20 96W	10.46	33.50	0.9954
Continental	Premium Contact 5 MO	225/55R17 97Y	4.95	36.22	0.9947
Continental	Premium Contact 6 MO	225/45R18 95Y	10.67	33.18	0.9926
Continental	Ultra Contact	215/60R17 96H	8.66	33.11	0.9933
Continental	4x4 Contact	235/50R18 101H	6.54	35.26	0.9944
Fortuna	Ecoplus UHP	235/45R18 98W	3.49	36.39	0.9968
Goodyear	Eagle F1 Asymmetric5 MO	225/40R19 93Y	15.10	30.31	0.9942
Goodyear	Efficient Grip	255/40R18 95Y	6.00	36.28	0.9963
Goodyear	Excellence	245/45R19 98Y	2.92	37.75	0.9969
Michelin	Pilot Sport4 SUV MO	255/45R20 105W	9.68	33.16	0.9943
Michelin	Primacy4 MO	225/45R18 95Y	5.97	34.32	0.9917
Nexen	N Fera Sport SU2 MO	225/55R18 102Y	8.56	34.03	0.9940
Pirelli	P Zero Nero GT	235/35R19 91Y	7.38	34.77	0.9950
Pirelli	Cinturato P7 MO	235/55R18 104T	8.00	33.97	0.9909
Pirelli	Scorpion Verde MO	235/55R19 101V	11.70	31.65	0.9957
Viking	ProTech NewGen	225/50R17 98Y	8.52	34.13	0.9899
Vredestein	Ultrac MO	225/55R18 102Y	2.41	36.07	0.9953
Hankook	Ventus S1 Evo 3 MO	225/55R18 102Y	13.26	31.03	0.9951
Dunlop	Dueler H/P Sport MO	235/55R19 101V	10.20	32.58	0.9922
Bridgestone	Potenza S001 MO	245/45R19 102Y	9.30	32.63	0.9918
Bridgestone	Turanza T005 MO-S	255/40R20 101Y	8.15	33.73	0.9956
Continental	ContiSportContact 5 MO1	255/50R19 103Y	9.96	34.14	0.9971
Continental	EcoContact 6Q MO	235/55R19 105W	6.67	33.66	0.9967
Cooper	Zeon Crossrange MO	255/45R20 105H	4.22	35.41	0.9909

continue on the next page

Manufacturer	Profile	Dimensions	a	b	R^2
Dunlop	SP Sport Maxx GT	275/30R20 97Y	9.72	33.29	0.9962
Goodyear	Eagle F1 SuperSport	265/35R20 99Y	7.89	35.49	0.9949
Kumho	Ecsta PS91 MO	245/45R18 100Y	10.33	32.47	0.9878
Michelin	ePrimacy MO	245/45R19 102Y	9.35	32.50	0.9933
Michelin	Latitude Sport 3 MO	235/60R18 103V	8.31	33.44	0.9949
Michelin	Pilot Sport 4 SUV MO	235/45R21 101Y	8.53	34.02	0.9967
Michelin	Pilot Sport 5 MO	245/40R19 98Y	6.38	35.60	0.9974
Michelin	Pilot Sport Cup 2 MO1	265/35R19 98Y	9.33	33.93	0.9952
Michelin	Pilot Sport 4 SUV MO	255/40R21 102Y	8.97	33.82	0.9956
Pirelli	P Zero (PZ4) MO	235/45R20 100T	6.77	34.84	0.9935
Pirelli	Powergy MO	255/45R19 104Y	11.43	32.21	0.9943
Pirelli	P Zero Corsa Asimmetrico2	265/30R19 93Y	6.31	35.98	0.9946
Dunlop	SP Sport Maxx GT MOE	245/35R20 95Y	10.31	32.87	0.9934
Yokohama	Advan Sport (V105) MO	235/60 R18 103V	9.98	32.66	0.9945
Continental	Premium Contact 6 MO-S	285/45R22 114Y	-0.22	38.83	0.9932

Table A.6: P-values of features in linear and lasso regression.

Feature	Linear Regression	Lasso Regression
Vehicle speed	0	0
Tire nominal width	0.0000007367	0.0032227598
Tire temperature shoulder	0.0134476673	0.6447870438
Tire aspect ratio	0.0660130903	0.9875592583
Air pressure	0.0000031891	0.0000041675
Shore hardness block	0.0395834399	0.8858826724
Tire profile depth	0.0014254838	0.0126329030
Tire speed index	0.1444596024	0.5512727475
Rim diameter	0.7699015670	0.9071323847
Wind direction	0.6017873573	0.9868901362
Wind speed	0.3822990259	0.8776349902
Air temperature	0.4479964555	0.2830135890
Humidity	0.2357968483	0.9948160756
Shore hardness shoulder	0.2476162282	0.5510819674
Road temperature	0.0159929044	0.1332388569
Tire temperature block	0.0305054756	1
Tire load index	0.0005953554	0.1063855711

Table A.7: Hyperparameter Optimization Importances for Test and Train Datasets of ANN.

Model	Split 1		Split 2		Split 3	
	Test	Train	Test	Train	Test	Train
Activation output	0.52	0.49	0.46	0.49	0.47	0.47
Activation hidden	0.24	0.26	0.29	0.30	0.22	0.28
Num Layers	0.02	0.03	0.03	0.03	0.03	0.05
Batch Size	0.03	0.03	0.03	0.02	0.04	0.02
Learning Rate	0.07	0.08	0.07	0.05	0.07	0.08
Dropout	0.08	0.06	0.08	0.07	0.14	0.13
Activation embedding	0.03	0.03	0.07	0.05	0.02	0.02
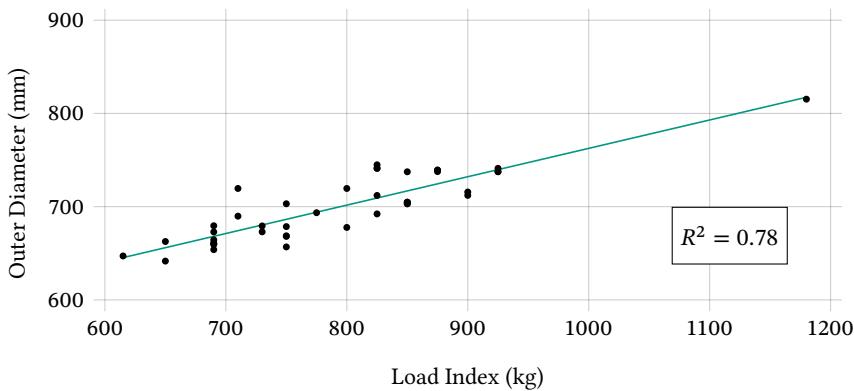
Figure A.1: Correlation of Load Index and Outer Diameter for the 42 different tires.
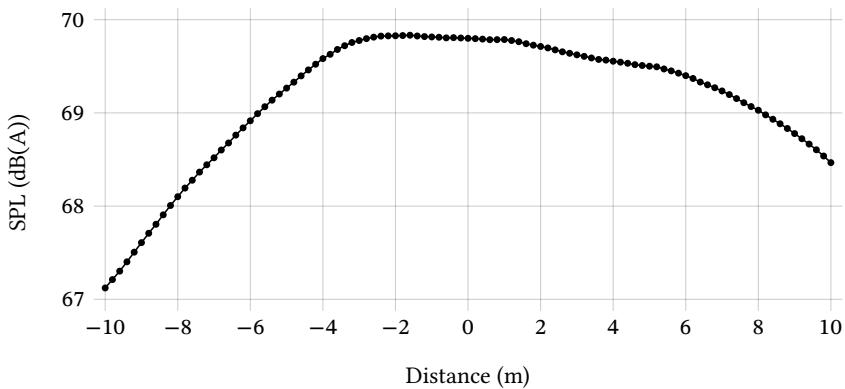
Figure A.2: Mean SPL of all 1443 measurements over distance.

B Additional Covariance Maps
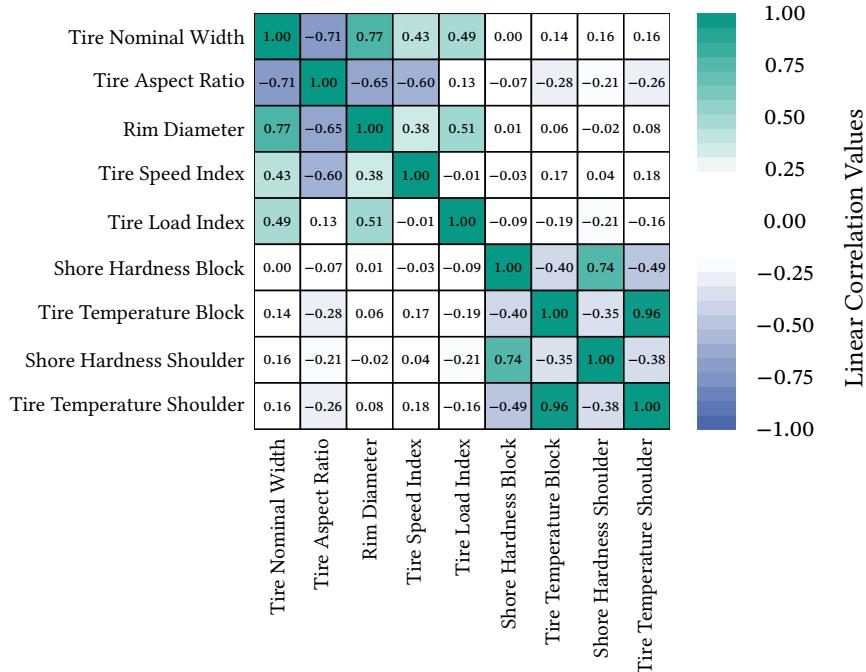
Figure B.1: Feature covariance as heat map for tire parameters for merged data set as used in CM.
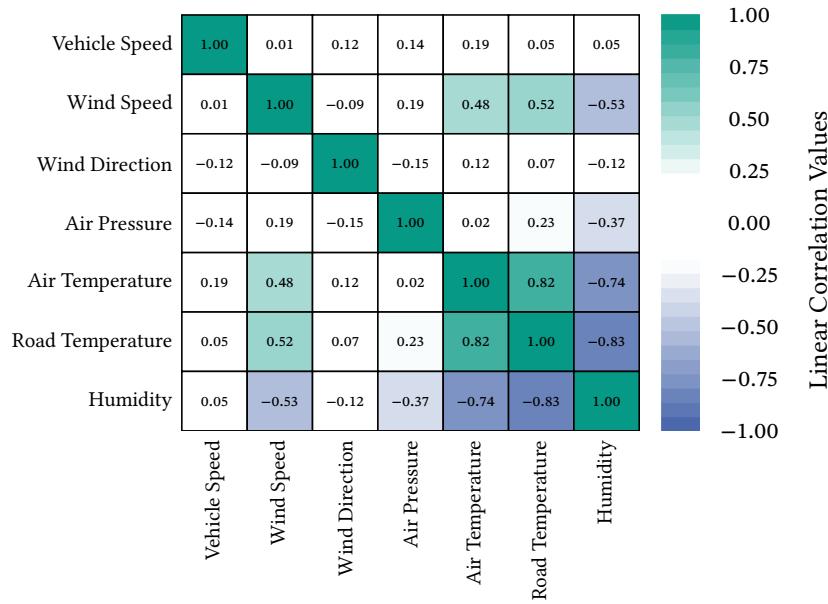
Figure B.2: Feature covariance as heat map for operation and environmental conditions for merged data set as used in CM.
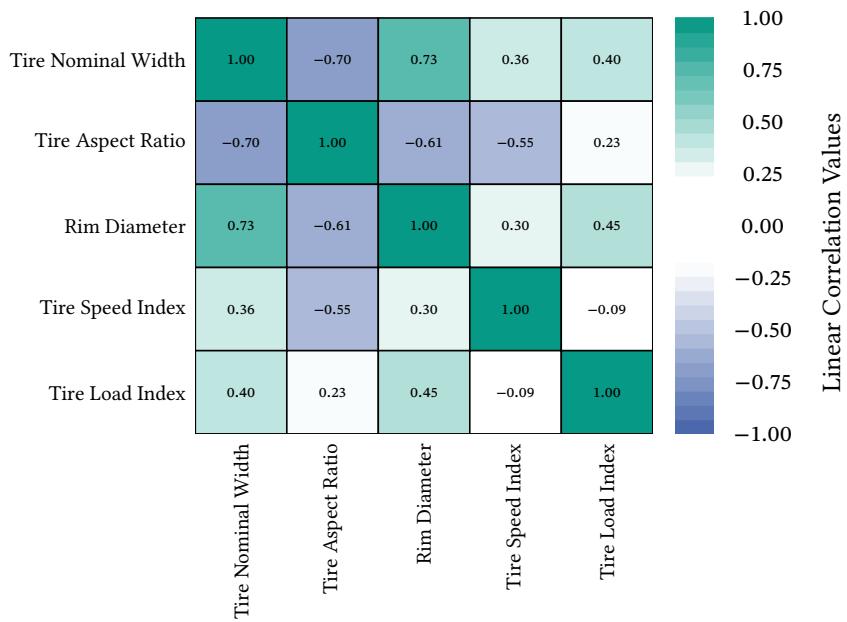
Figure B.3: Feature covariance as heat map for tire parameters of 42 different tires.

C Additional Model Performance CM$^{PP'}$

(a) Data split 0

(b) Data split 1

(c) Data split 2

(d) Data split 3

(e) Data split 4

(f) Data split 5

(g) Data split 6

(h) Data split 7

(i) Data split 8

(j) Data split 9

(k) Data split 10

(l) Data split 11

(m) Data split 12

(n) Data split 13

(o) Data split 14

(p) Data split 15

(q) Data split 16

(r) Data split 17

(s) Data split 18

(t) Data split 19
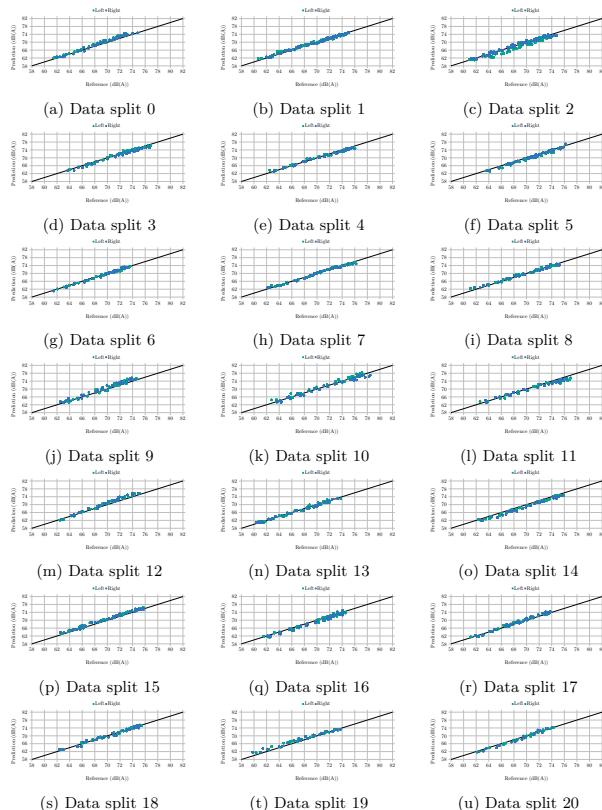
(u) Data split 20

Figure C.1: Prediction over reference values different data splits of CM$^{PP'}$ for test data.

(a) Data split 21

(b) Data split 22

(c) Data split 23

(d) Data split 24

(e) Data split 25

(f) Data split 26

(g) Data split 27

(h) Data split 28

(i) Data split 29

(j) Data split 30

(k) Data split 31

(l) Data split 32

(m) Data split 33

(n) Data split 34

(o) Data split 35

(p) Data split 36

(q) Data split 37

(r) Data split 38

(a) Data split 39

(b) Data split 40
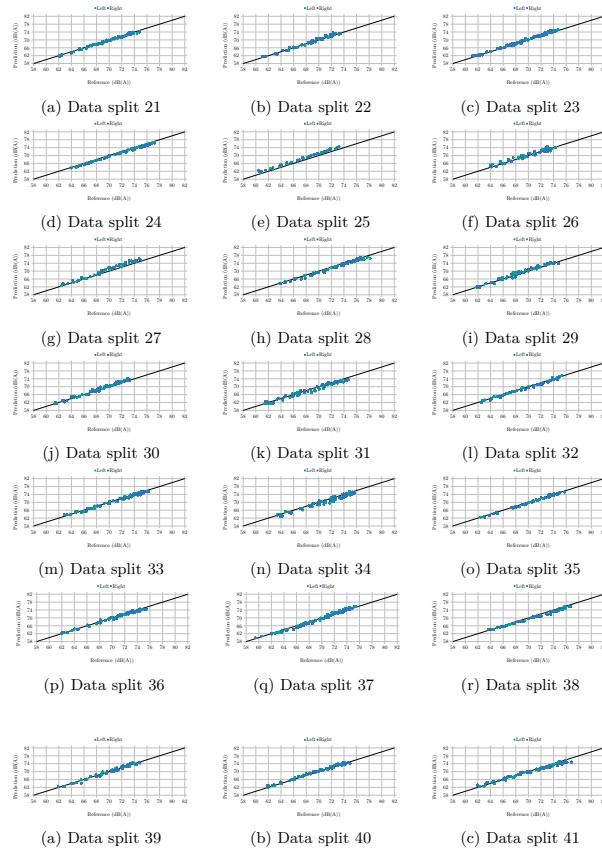
(c) Data split 41

Figure C.2: Prediction over reference values different data splits of CM^PP' for test data.

D Complementary Shapley Values for CNN
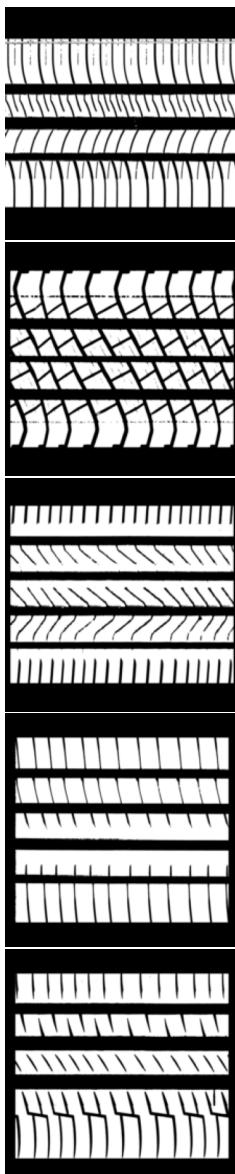
(a) inverted images   **(b)** left microphone   **(c)** right microphone
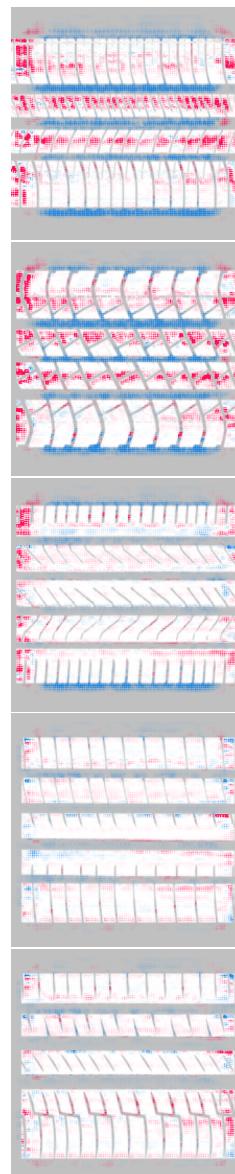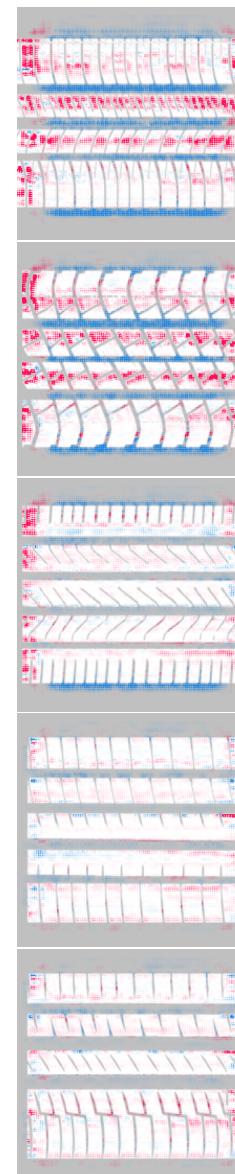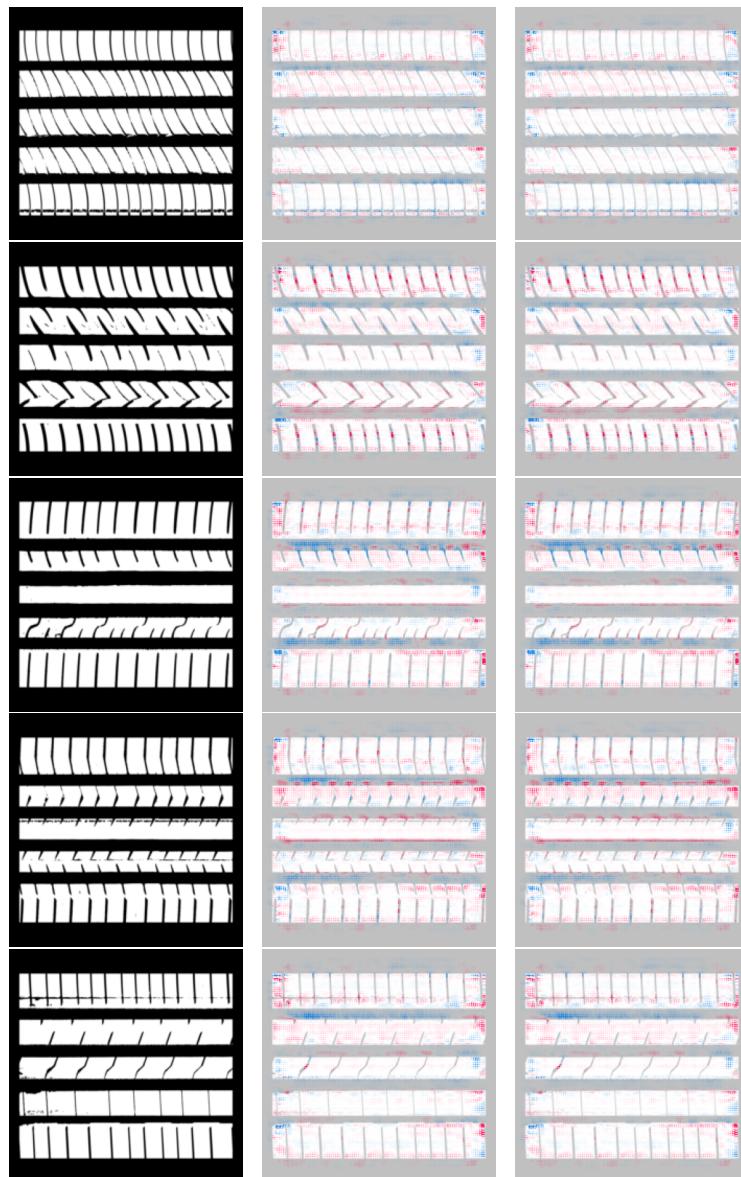
(a) inverted images   (b) left microphone   (c) right microphone

(a) inverted images   (b) left microphone   (c) right microphone

(a) inverted images   (b) left microphone   (c) right microphone
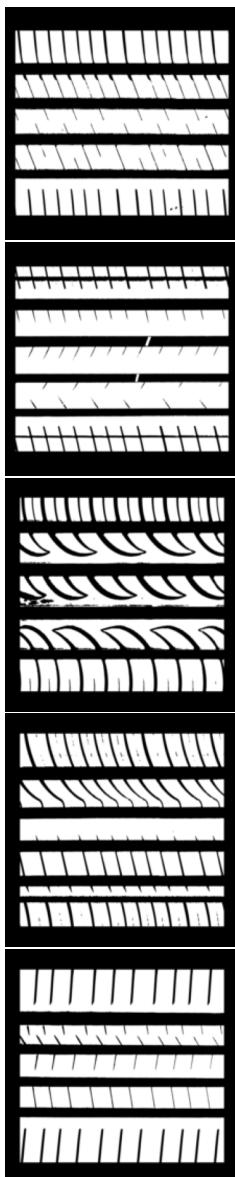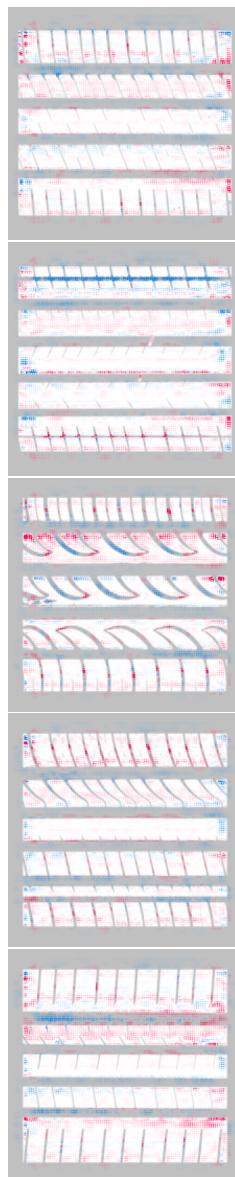
(a) inverted images   (b) left microphone   (c) right microphone
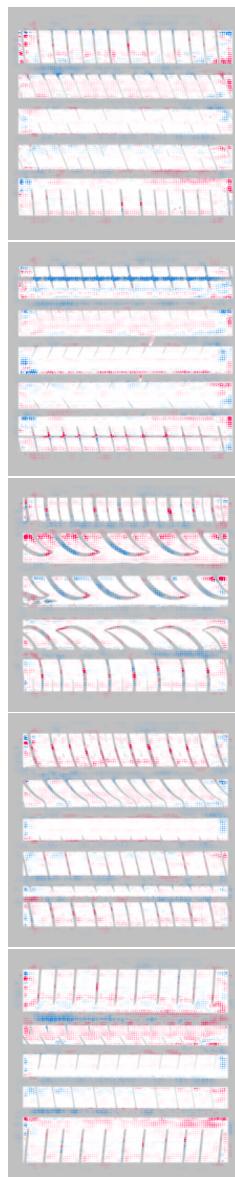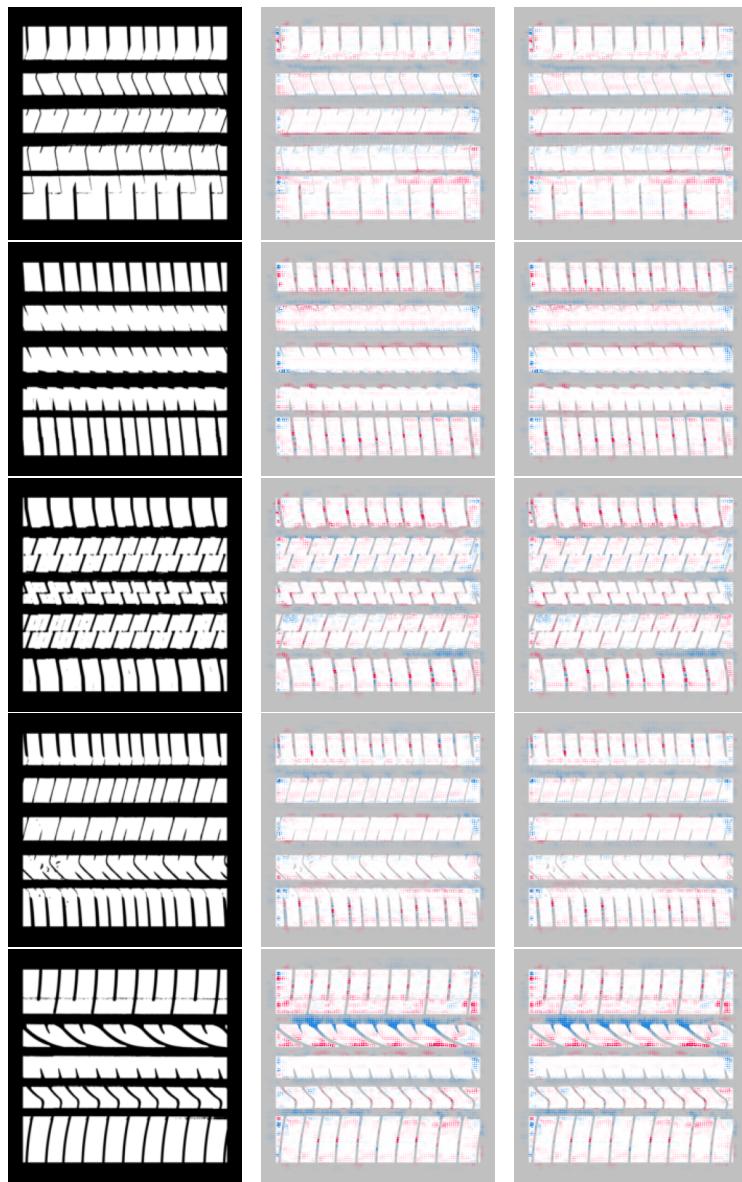
(a) inverted images   (b) left microphone   (c) right microphone
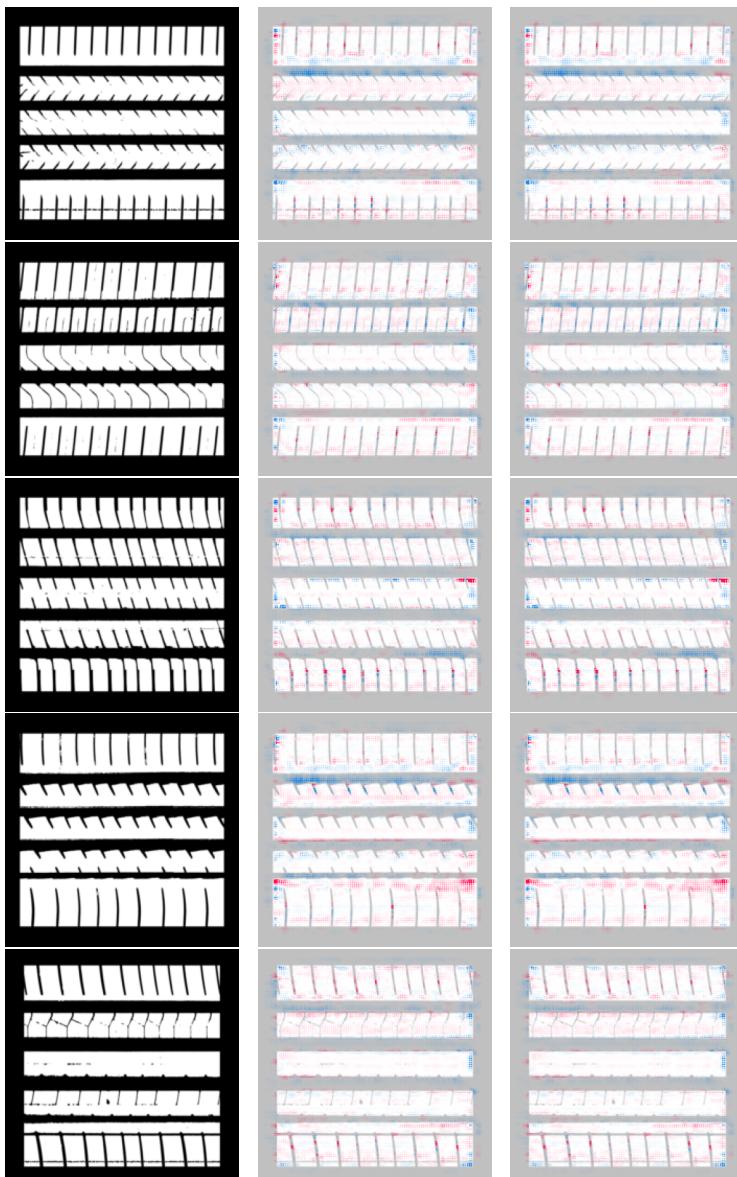
(a) inverted images   (b) left microphone   (c) right microphone
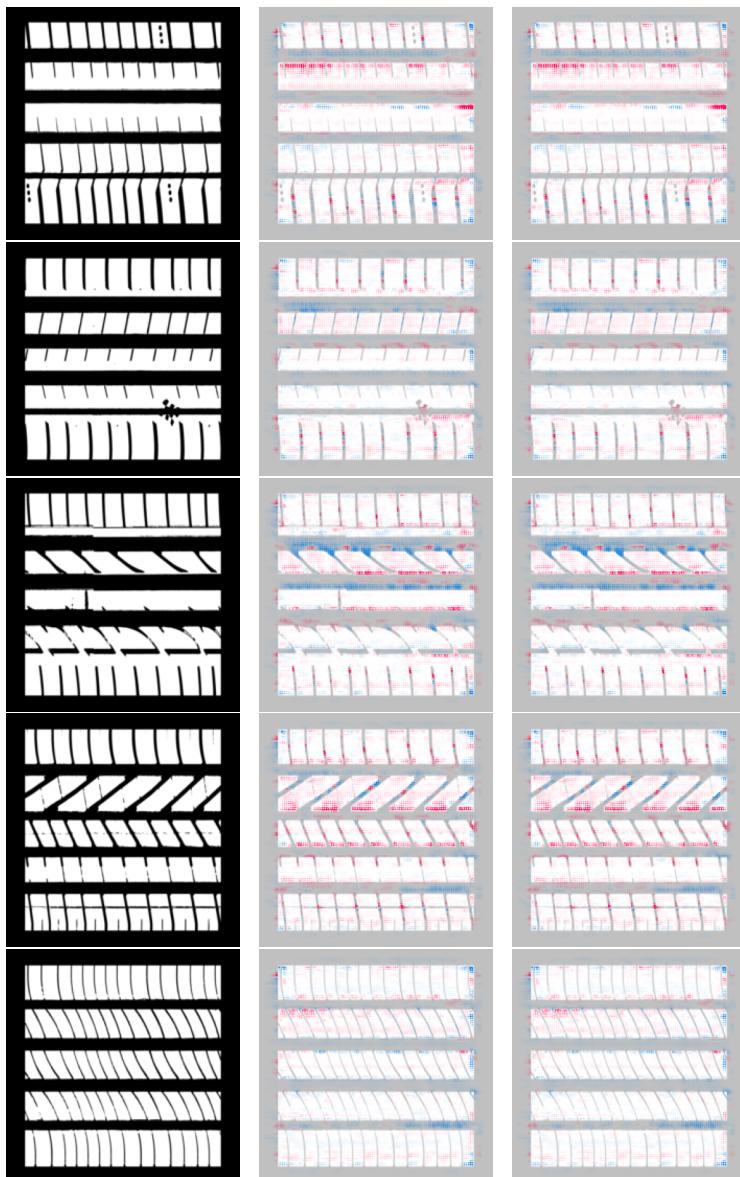
(a) inverted images   **(b)** left microphone   **(c)** right microphone
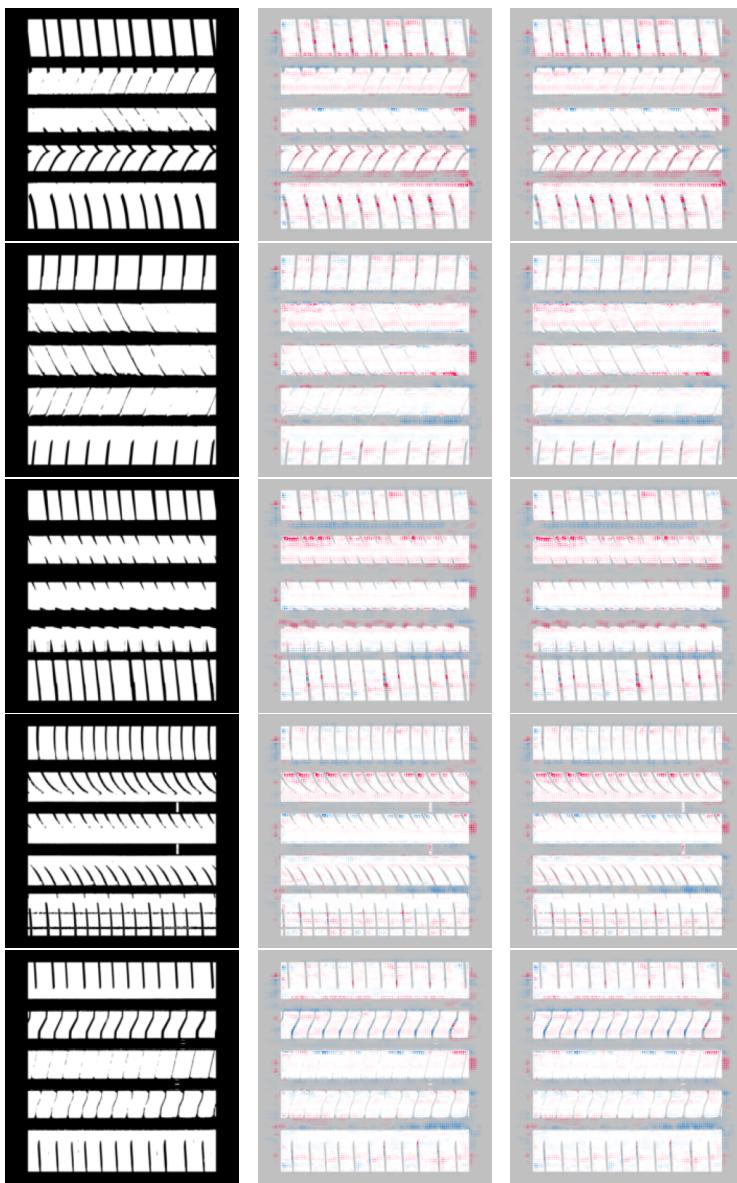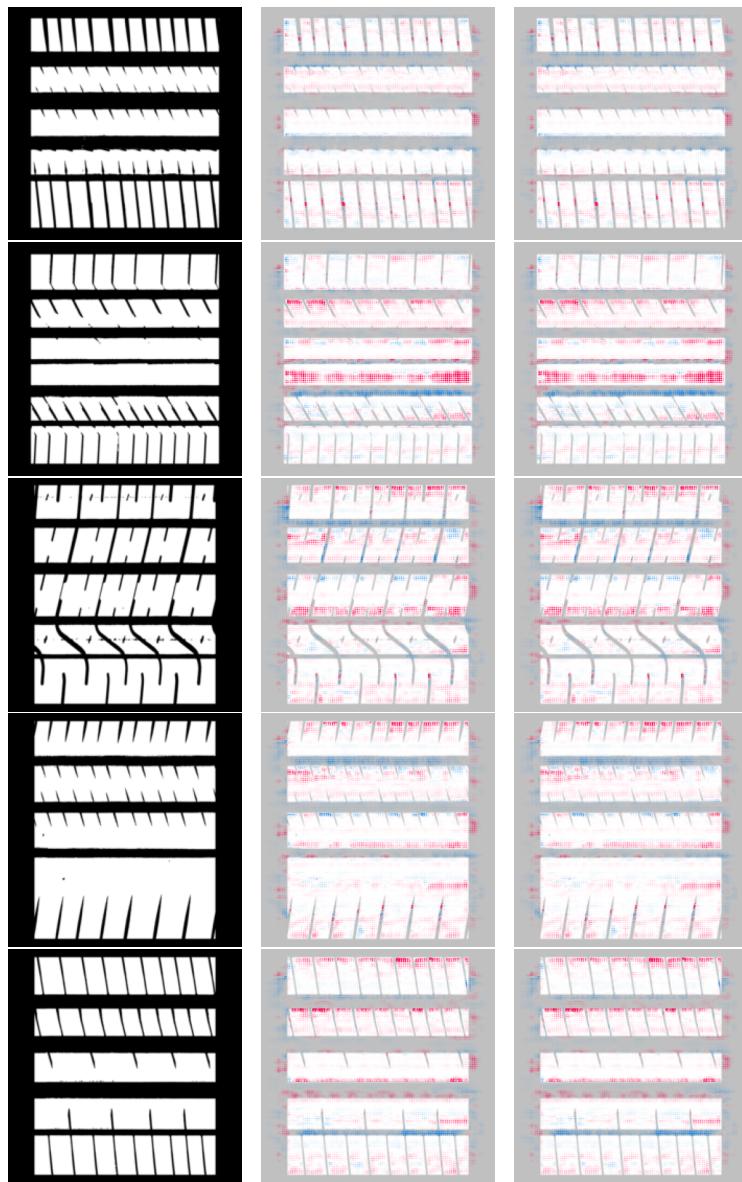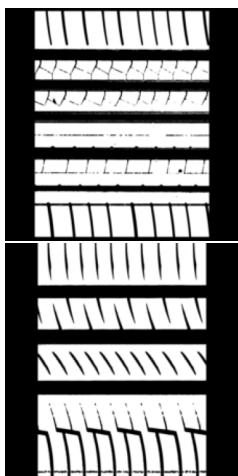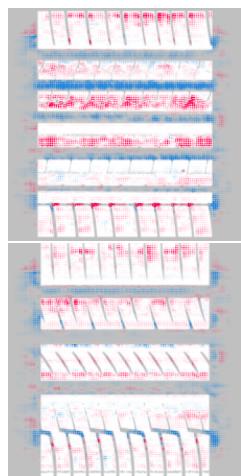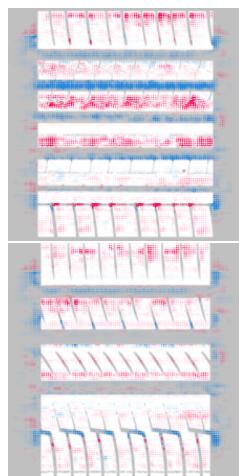
(a) inverted images   **(b)** left microphone   **(c)** right microphone

E Additional Images of Integrated Gradients

(a) Tire 0 semi slick left

(b) Tire 0 semi slick right

(c) Tire 0 slick left

(d) Tire 0 slick right

(e) Tire 1 semi slick left

(f) Tire 1 semi slick right

(g) Tire 1 slick left

(h) Tire 1 slick right

(i) Tire 2 semi slick left

(j) Tire 2 semi slick right

(k) Tire 2 slick left

(l) Tire 2 slick right

(m) Tire 3 semi slick left

(n) Tire 3 semi slick right

(o) Tire 3 slick left

(p) Tire 3 slick right

(q) Tire 4 semi slick left

(r) Tire 4 semi slick right

(s) Tire 4 slick left

(t) Tire 4 slick right

Figure E.1: Additional images of IG investigation for $CM_{\text{tire aspect ratio}}^{\text{Distance}}$.

(a) Tire 5 semi slick left

(b) Tire 5 semi slick right

(c) Tire 5 slick left

(d) Tire 5 slick right

(e) Tire 6 semi slick left

(f) Tire 6 semi slick right

(g) Tire 6 slick left

(h) Tire 6 slick right

(i) Tire 7 semi slick left

(j) Tire 7 semi slick right

(k) Tire 7 slick left

(l) Tire 7 slick right

(m) Tire 8 semi slick left

(n) Tire 8 semi slick right

(o) Tire 8 slick left

(p) Tire 8 slick right

(q) Tire 9 semi slick left

(r) Tire 9 semi slick right

(s) Tire 9 slick left

(t) Tire 9 slick right

(u) Tire 10 semi slick left

(v) Tire 10 semi slick right

(w) Tire 10 slick left

(x) Tire 10 slick right

Figure E.2: Additional images of IG investigation for $CM_{\text{tire aspect ratio}}^{\text{Distance}}$.

223

(a) Tire 11 semi slick left

(b) Tire 11 semi slick right

(c) Tire 11 slick left

(d) Tire 11 slick right

(e) Tire 12 semi slick left

(f) Tire 12 semi slick right

(g) Tire 12 slick left

(h) Tire 12 slick right

(i) Tire 13 semi slick left

(j) Tire 13 semi slick right

(k) Tire 13 slick left

(l) Tire 13 slick right

(m) Tire 14 semi slick left

(n) Tire 14 semi slick right

(o) Tire 14 slick left

(p) Tire 14 slick right

(q) Tire 15 semi slick left

(r) Tire 15 semi slick right

(s) Tire 15 slick left

(t) Tire 15 slick right

(u) Tire 16 semi slick left

(v) Tire 16 semi slick right

(w) Tire 16 slick left

(x) Tire 16 slick right

Figure E.3: Additional images of IG investigation for $CM_{\text{tire aspect ratio}}^{\text{Distance}}$.

(a) Tire 17 semi slick left

(b) Tire 17 semi slick right

(c) Tire 17 slick left

(d) Tire 17 slick right

(e) Tire 18 semi slick left

(f) Tire 18 semi slick right

(g) Tire 18 slick left

(h) Tire 18 slick right

(i) Tire 19 semi slick left

(j) Tire 19 semi slick right

(k) Tire 19 slick left

(l) Tire 19 slick right

(m) Tire 20 semi slick left

(n) Tire 20 semi slick right

(o) Tire 20 slick left

(p) Tire 20 slick right

(q) Tire 21 semi slick left

(r) Tire 21 semi slick right

(s) Tire 21 slick left

(t) Tire 21 slick right

(u) Tire 22 semi slick left

(v) Tire 22 semi slick right

(w) Tire 22 slick left

(x) Tire 22 slick right

Figure E.4: Additional images of IG investigation for $CM_{\text{tire aspect ratio}}^{\text{Distance}}$.

225

(a) Tire 23 semi slick left

(b) Tire 23 semi slick right

(c) Tire 23 slick left

(d) Tire 23 slick right

(e) Tire 24 semi slick left

(f) Tire 24 semi slick right

(g) Tire 24 slick left

(h) Tire 24 slick right

(i) Tire 25 semi slick left

(j) Tire 25 semi slick right

(k) Tire 25 slick left

(l) Tire 25 slick right

(m) Tire 26 semi slick left

(n) Tire 26 semi slick right

(o) Tire 26 slick left

(p) Tire 26 slick right

(q) Tire 27 semi slick left

(r) Tire 27 semi slick right

(s) Tire 27 slick left

(t) Tire 27 slick right

(u) Tire 28 semi slick left

(v) Tire 28 semi slick right

(w) Tire 28 slick left

(x) Tire 28 slick right

Figure E.5: Additional images of IG investigation for $CM_{\text{tire aspect ratio}}^{\text{Distance}}$.

(a) Tire 29 semi slick left

(b) Tire 29 semi slick right

(c) Tire 29 slick left

(d) Tire 29 slick right

(e) Tire 30 semi slick left

(f) Tire 30 semi slick right

(g) Tire 30 slick left

(h) Tire 30 slick right

(i) Tire 31 semi slick left

(j) Tire 31 semi slick right

(k) Tire 31 slick left

(l) Tire 31 slick right

(m) Tire 32 semi slick left

(n) Tire 32 semi slick right

(o) Tire 32 slick left

(p) Tire 32 slick right

(q) Tire 33 semi slick left

(r) Tire 33 semi slick right

(s) Tire 33 slick left

(t) Tire 33 slick right

(u) Tire 34 semi slick left

(v) Tire 34 semi slick right

(w) Tire 34 slick left

(x) Tire 34 slick right

Figure E.6: Additional images of IG investigation for $\text{CM}_{\text{tire aspect ratio}}^{\text{Distance}}$.

227

(a) Tire 35 semi slick left

(b) Tire 35 semi slick right

(c) Tire 35 slick left

(d) Tire 35 slick right

(e) Tire 36 semi slick left

(f) Tire 36 semi slick right

(g) Tire 36 slick left

(h) Tire 36 slick right

(i) Tire 37 semi slick left

(j) Tire 37 semi slick right

(k) Tire 37 slick left

(l) Tire 37 slick right

(m) Tire 38 semi slick left

(n) Tire 38 semi slick right

(o) Tire 38 slick left

(p) Tire 38 slick right

(q) Tire 39 semi slick left

(r) Tire 39 semi slick right

(s) Tire 39 slick left

(t) Tire 39 slick right

(u) Tire 40 semi slick left

(v) Tire 40 semi slick right

(w) Tire 40 slick left

(x) Tire 40 slick right

Figure E.7: Additional images of IG investigation for $CM_{\text{tire aspect ratio}}^{\text{Distance}}$.

(a) Tire 0 semi slick left

(b) Tire 0 semi slick right

(c) Tire 0 slick left

(d) Tire 0 slick right

(e) Tire 1 semi slick left

(f) Tire 1 semi slick right

(g) Tire 1 slick left

(h) Tire 1 slick right

(i) Tire 2 semi slick left

(j) Tire 2 semi slick right

(k) Tire 2 slick left

(l) Tire 2 slick right

(m) Tire 3 semi slick left

(n) Tire 3 semi slick right

(o) Tire 3 slick left

(p) Tire 3 slick right

(q) Tire 4 semi slick left

(r) Tire 4 semi slick right

(s) Tire 4 slick left

(t) Tire 4 slick right

Figure E.8: Additional images of IG investigation for $CM_{\text{rim diameter}}^{\text{Distance}}$.

(a) Tire 5 semi slick left

(b) Tire 5 semi slick right

(c) Tire 5 slick left

(d) Tire 5 slick right

(e) Tire 6 semi slick left

(f) Tire 6 semi slick right

(g) Tire 6 slick left

(h) Tire 6 slick right

(i) Tire 7 semi slick left

(j) Tire 7 semi slick right

(k) Tire 7 slick left

(l) Tire 7 slick right

(m) Tire 8 semi slick left

(n) Tire 8 semi slick right

(o) Tire 8 slick left

(p) Tire 8 slick right

(q) Tire 9 semi slick left

(r) Tire 9 semi slick right

(s) Tire 9 slick left

(t) Tire 9 slick right

(u) Tire 10 semi slick left

(v) Tire 10 semi slick right

(w) Tire 10 slick left

(x) Tire 10 slick right

Figure E.9: Additional images of IG investigation for $CM_{\text{rim diameter}}^{\text{Distance}}$.

(a) Tire 11 semi slick left

(b) Tire 11 semi slick right

(c) Tire 11 slick left

(d) Tire 11 slick right

(e) Tire 12 semi slick left

(f) Tire 12 semi slick right

(g) Tire 12 slick left

(h) Tire 12 slick right

(i) Tire 13 semi slick left

(j) Tire 13 semi slick right

(k) Tire 13 slick left

(l) Tire 13 slick right

(m) Tire 14 semi slick left

(n) Tire 14 semi slick right

(o) Tire 14 slick left

(p) Tire 14 slick right

(q) Tire 15 semi slick left

(r) Tire 15 semi slick right

(s) Tire 15 slick left

(t) Tire 15 slick right

(u) Tire 16 semi slick left

(v) Tire 16 semi slick right

(w) Tire 16 slick left

(x) Tire 16 slick right

Figure E.10: Additional images of IG investigation for $\mathrm{CM}_{\text{rim diameter}}^{\text{Distance}}$.

(a) Tire 17 semi slick left

(b) Tire 17 semi slick right

(c) Tire 17 slick left

(d) Tire 17 slick right

(e) Tire 18 semi slick left

(f) Tire 18 semi slick right

(g) Tire 18 slick left

(h) Tire 18 slick right

(i) Tire 19 semi slick left

(j) Tire 19 semi slick right

(k) Tire 19 slick left

(l) Tire 19 slick right

(m) Tire 20 semi slick left

(n) Tire 20 semi slick right

(o) Tire 20 slick left

(p) Tire 20 slick right

(q) Tire 21 semi slick left

(r) Tire 21 semi slick right

(s) Tire 21 slick left

(t) Tire 21 slick right

(u) Tire 22 semi slick left

(v) Tire 22 semi slick right

(w) Tire 22 slick left

(x) Tire 22 slick right

Figure E.11: Additional images of IG investigation for $CM_{\text{rim diameter}}^{\text{Distance}}$.

(a) Tire 23 semi slick left

(b) Tire 23 semi slick right

(c) Tire 23 slick left

(d) Tire 23 slick right

(e) Tire 24 semi slick left

(f) Tire 24 semi slick right

(g) Tire 24 slick left

(h) Tire 24 slick right

(i) Tire 25 semi slick left

(j) Tire 25 semi slick right

(k) Tire 25 slick left

(l) Tire 25 slick right

(m) Tire 26 semi slick left

(n) Tire 26 semi slick right

(o) Tire 26 slick left

(p) Tire 26 slick right

(q) Tire 27 semi slick left

(r) Tire 27 semi slick right

(s) Tire 27 slick left

(t) Tire 27 slick right

(u) Tire 28 semi slick left

(v) Tire 28 semi slick right

(w) Tire 28 slick left

(x) Tire 28 slick right

233

Figure E.12: Additional images of IG investigation for $CM_{rim\ diameter}^{Distance}$.

(a) Tire 29 semi slick left

(b) Tire 29 semi slick right

(c) Tire 29 slick left

(d) Tire 29 slick right

(e) Tire 30 semi slick left

(f) Tire 30 semi slick right

(g) Tire 30 slick left

(h) Tire 30 slick right

(i) Tire 31 semi slick left

(j) Tire 31 semi slick right

(k) Tire 31 slick left

(l) Tire 31 slick right

(m) Tire 32 semi slick left

(n) Tire 32 semi slick right

(o) Tire 32 slick left

(p) Tire 32 slick right

(q) Tire 33 semi slick left

(r) Tire 33 semi slick right

(s) Tire 33 slick left

(t) Tire 33 slick right

(u) Tire 34 semi slick left

(v) Tire 34 semi slick right

(w) Tire 34 slick left

(x) Tire 34 slick right

Figure E.13: Additional images of IG investigation for $\text{CM}_{\text{rim diameter}}^{\text{Distance}}$.

(a) Tire 35 semi slick left

(b) Tire 35 semi slick right

(c) Tire 35 slick left

(d) Tire 35 slick right

(e) Tire 36 semi slick left

(f) Tire 36 semi slick right

(g) Tire 36 slick left

(h) Tire 36 slick right

(i) Tire 37 semi slick left

(j) Tire 37 semi slick right

(k) Tire 37 slick left

(l) Tire 37 slick right

(m) Tire 38 semi slick left

(n) Tire 38 semi slick right

(o) Tire 38 slick left

(p) Tire 38 slick right

(q) Tire 39 semi slick left

(r) Tire 39 semi slick right

(s) Tire 39 slick left

(t) Tire 39 slick right

(u) Tire 40 semi slick left

(v) Tire 40 semi slick right



(w) Tire 40 slick left

(x) Tire 40 slick right

235

Figure E.14: Additional images of IG investigation for $CM_{\text{rim diameter}}^{\text{Distance}}$.

(a) Tire 0 semi slick left

(b) Tire 0 semi slick right

(c) Tire 0 slick left

(d) Tire 0 slick right

(e) Tire 1 semi slick left

(f) Tire 1 semi slick right

(g) Tire 1 slick left

(h) Tire 1 slick right

(i) Tire 2 semi slick left

(j) Tire 2 semi slick right

(k) Tire 2 slick left

(l) Tire 2 slick right

(m) Tire 3 semi slick left

(n) Tire 3 semi slick right

(o) Tire 3 slick left

(p) Tire 3 slick right

(q) Tire 4 semi slick left

(r) Tire 4 semi slick right

(s) Tire 4 slick left

(t) Tire 4 slick right

Figure E.15: Additional images of IG investigation for $CM_{outer\ diameter}^{Distance}$.

(a) Tire 5 semi slick left

(b) Tire 5 semi slick right

(c) Tire 5 slick left

(d) Tire 5 slick right

(e) Tire 6 semi slick left

(f) Tire 6 semi slick right

(g) Tire 6 slick left

(h) Tire 6 slick right

(i) Tire 7 semi slick left

(j) Tire 7 semi slick right

(k) Tire 7 slick left

(l) Tire 7 slick right

(m) Tire 8 semi slick left

(n) Tire 8 semi slick right

(o) Tire 8 slick left

(p) Tire 8 slick right

(q) Tire 9 semi slick left

(r) Tire 9 semi slick right

(s) Tire 9 slick left

(t) Tire 9 slick right

(u) Tire 10 semi slick left

(v) Tire 10 semi slick right

(w) Tire 10 slick left

(x) Tire 10 slick right

**Figure E.16:** Additional images of IG investigation for $CM_{outer\ diameter}^{Distance}$.

(a) Tire 11 semi slick left

(b) Tire 11 semi slick right

(c) Tire 11 slick left

(d) Tire 11 slick right

(e) Tire 12 semi slick left

(f) Tire 12 semi slick right

(g) Tire 12 slick left

(h) Tire 12 slick right

(i) Tire 13 semi slick left

(j) Tire 13 semi slick right

(k) Tire 13 slick left

(l) Tire 13 slick right

(m) Tire 14 semi slick left

(n) Tire 14 semi slick right

(o) Tire 14 slick left

(p) Tire 14 slick right

(q) Tire 15 semi slick left

(r) Tire 15 semi slick right

(s) Tire 15 slick left

(t) Tire 15 slick right

(u) Tire 16 semi slick left

(v) Tire 16 semi slick right

(w) Tire 16 slick left

(x) Tire 16 slick right

Figure E.17: Additional images of IG investigation for $CM_{\text{outer diameter}}^{\text{Distance}}$.

(a) Tire 17 semi slick left

(b) Tire 17 semi slick right

(c) Tire 17 slick left

(d) Tire 17 slick right

(e) Tire 18 semi slick left

(f) Tire 18 semi slick right

(g) Tire 18 slick left

(h) Tire 18 slick right

(i) Tire 19 semi slick left

(j) Tire 19 semi slick right

(k) Tire 19 slick left

(l) Tire 19 slick right

(m) Tire 20 semi slick left

(n) Tire 20 semi slick right

(o) Tire 20 slick left

(p) Tire 20 slick right

(q) Tire 21 semi slick left

(r) Tire 21 semi slick right

(s) Tire 21 slick left

(t) Tire 21 slick right

(u) Tire 22 semi slick left

(v) Tire 22 semi slick right

(w) Tire 22 slick left

(x) Tire 22 slick right

Figure E.18: Additional images of IG investigation for $CM_{\text{outer diameter}}^{\text{Distance}}$.

(a) Tire 23 semi slick left

(b) Tire 23 semi slick right

(c) Tire 23 slick left

(d) Tire 23 slick right

(e) Tire 24 semi slick left

(f) Tire 24 semi slick right

(g) Tire 24 slick left

(h) Tire 24 slick right

(i) Tire 25 semi slick left

(j) Tire 25 semi slick right

(k) Tire 25 slick left

(l) Tire 25 slick right

(m) Tire 26 semi slick left

(n) Tire 26 semi slick right

(o) Tire 26 slick left

(p) Tire 26 slick right

(q) Tire 27 semi slick left

(r) Tire 27 semi slick right

(s) Tire 27 slick left

(t) Tire 27 slick right

(u) Tire 28 semi slick left

(v) Tire 28 semi slick right

(w) Tire 28 slick left

(x) Tire 28 slick right

Figure E.19: Additional images of IG investigation for $CM_{\text{outer diameter}}^{\text{Distance}}$.

(a) Tire 29 semi slick left

(b) Tire 29 semi slick right

(c) Tire 29 slick left

(d) Tire 29 slick right

(e) Tire 30 semi slick left

(f) Tire 30 semi slick right

(g) Tire 30 slick left

(h) Tire 30 slick right

(i) Tire 31 semi slick left

(j) Tire 31 semi slick right

(k) Tire 31 slick left

(l) Tire 31 slick right

(m) Tire 32 semi slick left

(n) Tire 32 semi slick right

(o) Tire 32 slick left

(p) Tire 32 slick right

(q) Tire 33 semi slick left

(r) Tire 33 semi slick right

(s) Tire 33 slick left

(t) Tire 33 slick right

(u) Tire 34 semi slick left

(v) Tire 34 semi slick right

(w) Tire 34 slick left

(x) Tire 34 slick right

Figure E.20: Additional images of IG investigation for $CM_{\text{outer diameter}}^{\text{Distance}}$.

241

(a) Tire 35 semi slick left

(b) Tire 35 semi slick right

(c) Tire 35 slick left

(d) Tire 35 slick right

(e) Tire 36 semi slick left

(f) Tire 36 semi slick right

(g) Tire 36 slick left

(h) Tire 36 slick right

(i) Tire 37 semi slick left

(j) Tire 37 semi slick right

(k) Tire 37 slick left

(l) Tire 37 slick right

(m) Tire 38 semi slick left

(n) Tire 38 semi slick right

(o) Tire 38 slick left

(p) Tire 38 slick right

(q) Tire 39 semi slick left

(r) Tire 39 semi slick right

(s) Tire 39 slick left

(t) Tire 39 slick right

(u) Tire 40 semi slick left

(v) Tire 40 semi slick right

(w) Tire 40 slick left

(x) Tire 40 slick right

Figure E.21: Additional images of IG investigation for $CM_{\text{outer diameter}}^{\text{Distance}}$.

F   Additional Model Interpretation

F.1   Additional CM^{Distance} Interpretation

The values in this subsections are calculated for ten independently trained models and averaged for these. This step is taken to somewhat compensate the multicollinearities.
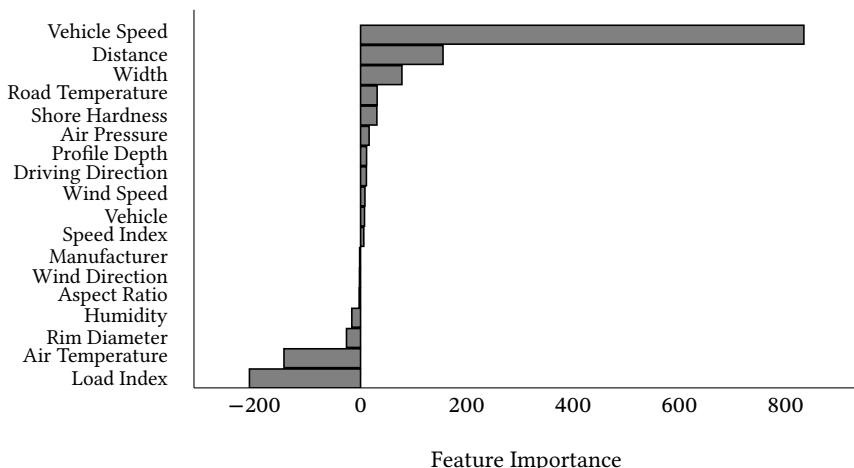
Figure F.1: Feature Importance of numerical input into CM distance model calculated through integrated gradients. Values averaged for ten independently trained models on whole data set, averaged for both microphone sides and slick and semi slick images.

(a) Driving Direction

(b) Tire Manufacturer

(c) Measurement Vehicle
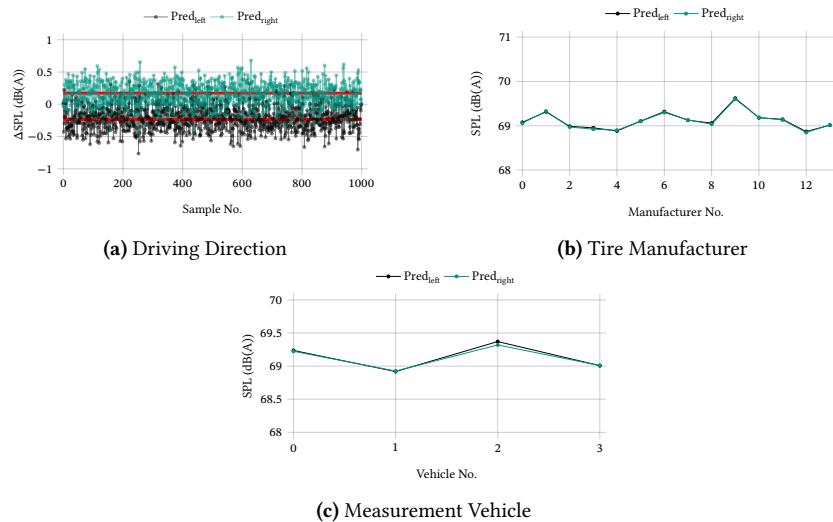
Figure F.2: Feature importance of one-hot encoded features of $CM^{Distance}$.

(a) Tire Width

(b) Tire Aspect Ratio

(c) Tire Speed Index

(d) Tire Load Index

(e) Rim Size

(f) Tire Profile Depth
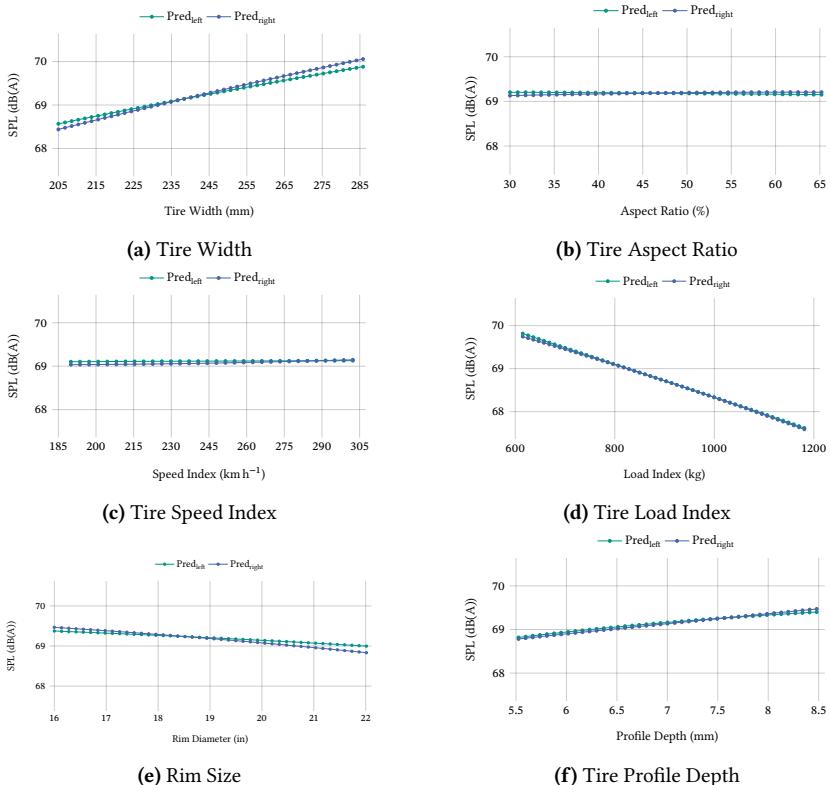
Figure F.3: Feature importance of CM$^{\text{Distance}}$

(a) Vehicle Speed

(b) Position on Test Track
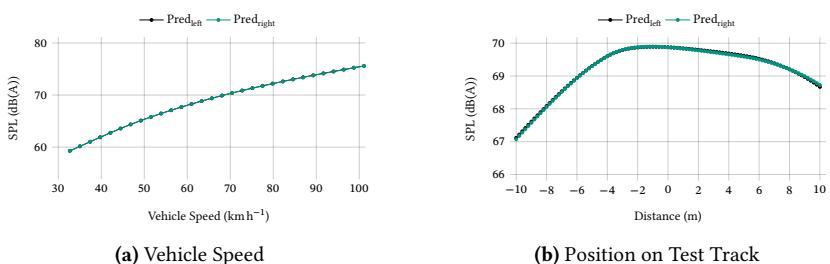
Figure F.4: Feature importance of CM$^{\text{Distance}}$.

(a) Air Temperature

(b) Road Temperature
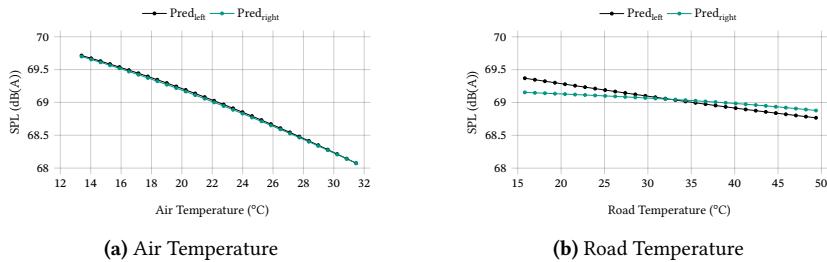
Figure F.5: Feature importance of $CM^{Distance}$.

F.2 Additional $CM^{Distance}_{aspect\ ratio}$ Interpretation

(a) Vehicle Speed

(b) Position on Test Track
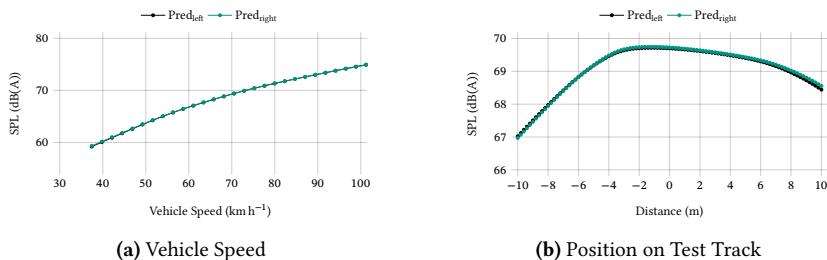
Figure F.6: Feature importance of $CM^{Distance}_{aspect\ ratio}$.

(a) Tire Nominal Width

(b) Rim Diameter

(c) Tire Speed Index

(d) Tire Load Index

(e) Tire Profile Depth
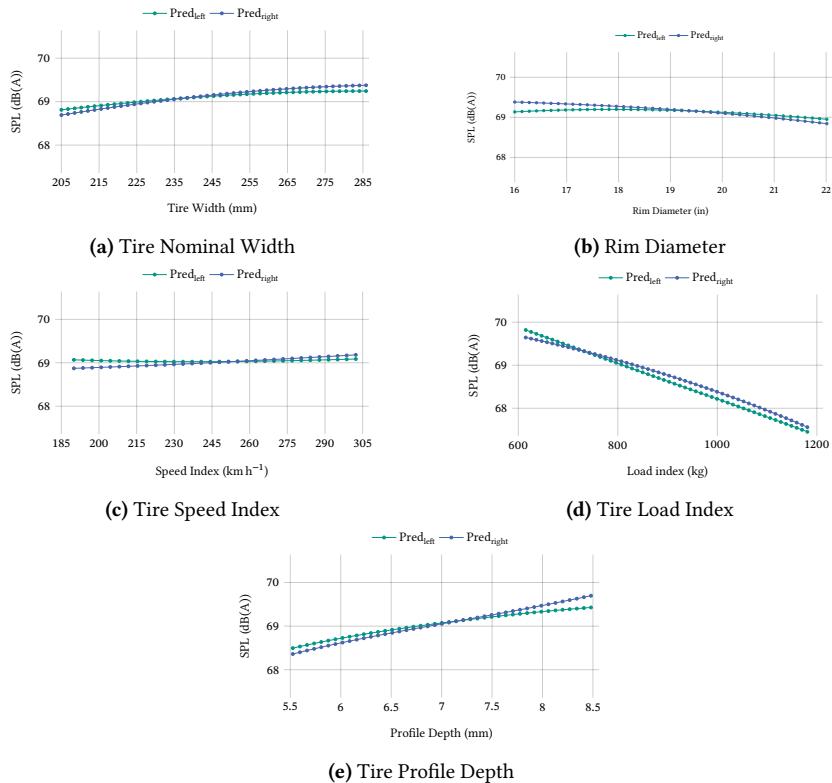
Figure F.7: Feature importance of CM$_{\text{aspect ratio}}^{\text{Distance}}$.

(a) Air Temperature

(b) Air Pressure
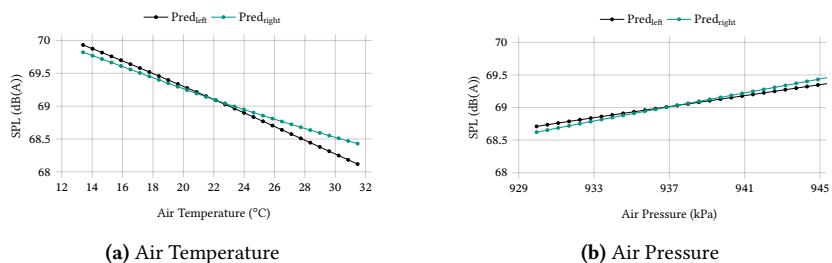
Figure F.8: Feature importance of CM$_{\text{aspect ratio}}^{\text{Distance}}$.

(a) Driving Direction

(b) Tire Manufacturer

(c) Measurement Vehicle
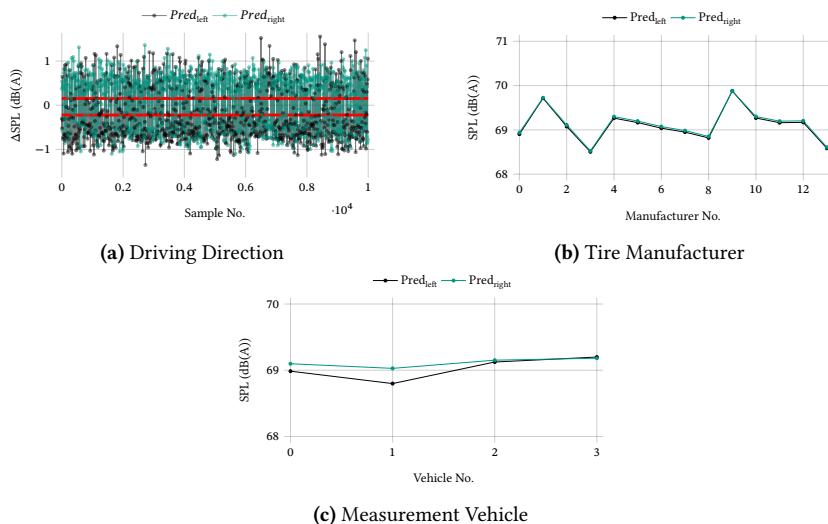
Figure F.9: Feature importance of one-hot encoded features of $CM_{aspect\ ratio}^{Distance}$.

F.3   Additional $CM_{outer\ diameter}^{Distance}$ Interpretation

(a) Vehicle Speed

(b) Position on Test Track
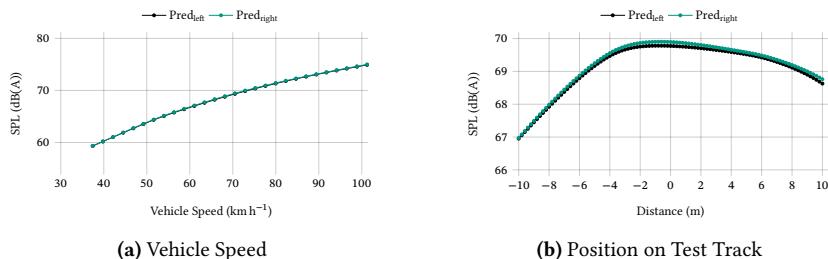
Figure F.10: Feature importance of $CM_{outer\ diameter}^{Distance}$.

(a) Tire Nominal Width

(b) Outer Diameter

**(c)** Tire Speed Index

(d) Tire Profile Depth
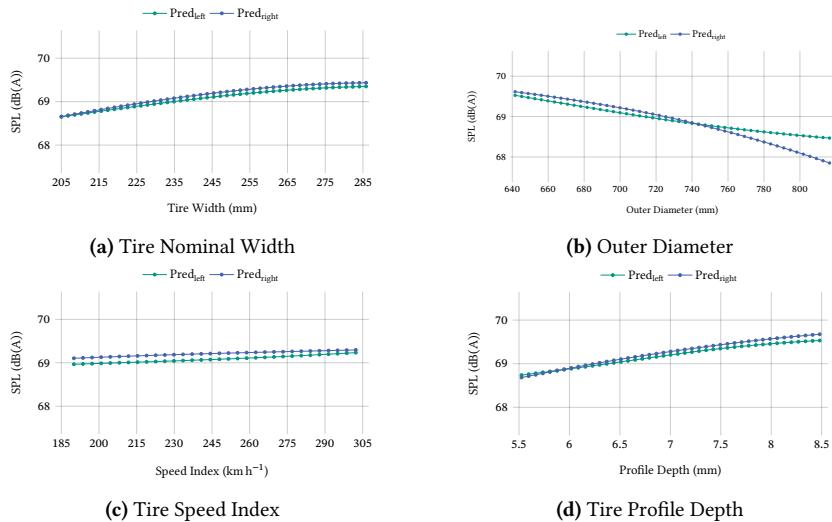
Figure F.11: Feature importance of CM$_{\text{outer diameter}}^{\text{Distance}}$

(a) Air Temperature

(b) Air Pressure
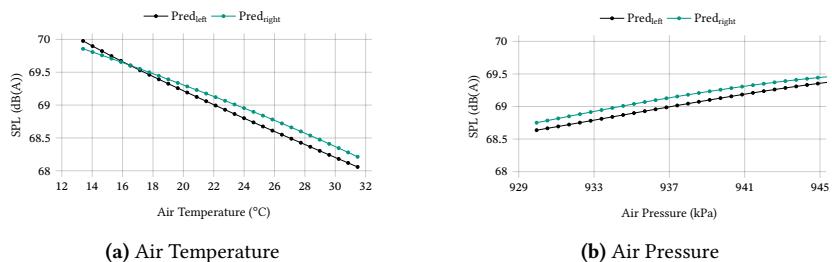
**Figure F.12:** Feature importance of CM$_{\text{outer diameter}}^{\text{Distance}}$.

(a) Driving Direction

(b) Tire Manufacturer

(c) Measurement Vehicle
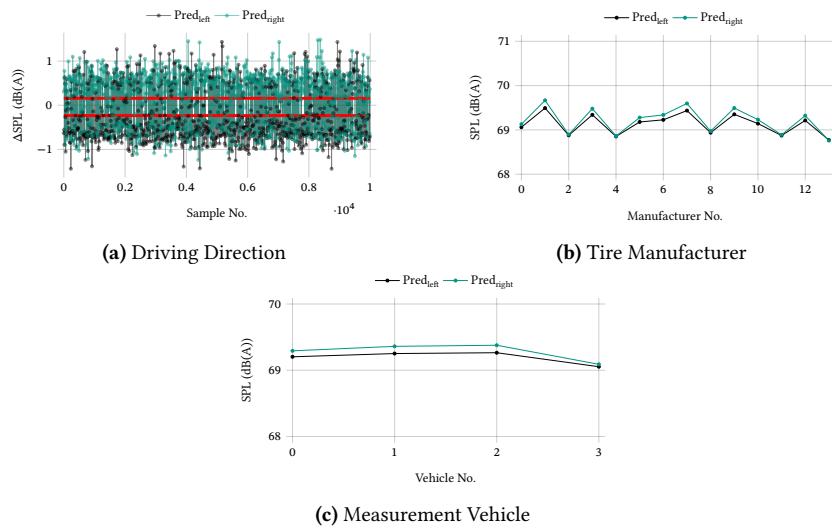
Figure F.13: Feature importance of one-hot encoded features of $CM^{Distance}_{outer\ diameter}$.