Tampere University

Aapo Hakala

SOUND BASED CLASSIFICATION OF STUDDED TIRES

Automatic Tire Classification System

Faculty of Information Technology and Communication Sciences
Bachelor's thesis
April 2020

ABSTRACT

Aapo Hakala: Sound Based Classification of Studded Tires
Bachelor's thesis
Tampere University
Bachelor's Degree Programme in Electrical Engineering
April 2020

The use of studded tires causes rutting of asphalt pavements and generates street dust to the environment. The maintenance of paved roads and cleaning of street dust requires resources and causes health risks. These effects are notable especially in spring time when the snow and ice has melted away from road surfaces. In order to predict these phenomena, the number of vehicles using studded tires should be measured continuously. Previously the estimations about the proportions of winter and summer tires have been created based on figures provided by car service companies that offer tire changing services. Occasional hearing based roadside sample surveys have also been made. Unlike the statistics from car service companies, hearing based data collection methods provide location and time specific information about the use of studded tires. Hearing based data collection is a difficult and labour-consuming task and it has not been applied widely. The purpose of this thesis was to find out if an automatic tire classification system could be implemented to collect data about the use of studded tires. A dataset of in-road audio recordings was exploited in the study. The dataset was collected from two measurement sites by using contact microphones under the road pavement. The measuring points were placed next to automatic traffic measurement stations that are used by Finnish Transport Infrastructure Agency in data collection purposes.

Digital signal processing and machine learning was applied in the designing of the tire classification system. A passenger car detector was implemented to restrict the classification only for tires of passenger cars and to determine the exact bypass times of detected vehicles. Feature extraction from the audio data was done according to modeling of the human auditory system. Two versions of the tire classifier were designed, one based on support vector machine and the other on multilayer perceptron. The dataset was annotated by labelling the recordings with the information about the vehicle class and the tire type used in the vehicle. The recordings of passenger cars were used in the training and testing of the classifier-models. The split of data into a training set and test set was done according to recording locations, meaning that data from one location was named as the training set while the remaining data from the other location was used as test set. This way the generalization of the system could be verified as the classifier-models could not learn the recording location-specific factors of the test set during the training. A comparison of the two classifier models was made according to the results of the experiments that were carried out with the test set.

The results of the experiments prove that automatic and instant tire classification is possible with the proposed methods. Both the passenger car detector and the tire classifier performed well in the experiments by scoring about 95% test accuracy. The differences between the results of the classifier models were small. The results imply that the system is able to generalize its knowledge from one recording environment to another without being explicitly trained to do so. However, due to the small amount of measurement sites used in the experiments, it is impossible to make reliable conclusions about general adaptivity of the system without further research. In order to improve the performance and reliability of the system, more data from new measurement sites should be collected in the follow-up research.

Keywords: tire classification, studded tires, automatic sound classification, traffic monitoring

TIIVISTELMÄ

Aapo Hakala: Ääneen perustuva nastarenkaiden luokittelu
Kandidaatintyö
Tampereen yliopisto
Sähkötekniikan tutkinto-ohjelma
Huhtikuu 2020

Nastarenkaiden käyttö aiheuttaa päällystettyjen tiepintojen urautumista ja katupölyn muodostumista tieympäristöön. Tiepintojen kulumisen seurauksena teiden kunnossapitoon ja katu-jen siistimiseen kuluu vuosittain paljon resursseja. Lisäksi katupölystä aiheutuu terveyshaittoja erityisesti kevätaikaan, jolloin teiden jää- ja lumipeitteiden sulaminen sekä tiepintojen kuivumi-nen nopeuttavat niiden kulumista. Ilmiöitä voitaisiin ennustaa paremmin, jos nastarenkaita käyttävien ajoneuvojen lukumäärää pystyttäisiin tilastoimaan. Talvi- ja kesärenkaiden tilastol-lista jakaumaa tieliikenteessä on aiemmin arvioitu renkaidenvaihtoa tarjoavien liikkeiden ilmoit-tamien tietojen perusteella. Myös yksittäisiä rengasäänien kuunteluun pohjautuvia tienvarsimit-tauksia on suoritettu kyselytutkimusten tueksi. Korvakuulolla tehtyjen tienvarsimittausten etu-na on niiden täsmälliset sijainti- ja aikatiedot, joita ei voida selvittää kyselytutkimuksilla. Tien-varsimittausten tekeminen manuaalisesti on kuitenkin työlästä ja haastavaa, joten laajamittais-ta kuuloon perustuvaa tiedonkeräystä ei olla harjoitettu. Tämän työn tarkoituksena oli selvittää, voidaanko henkilöautojen rengastyyppien automaattinen luokittelu toteuttaa rengasäänien mit-taamiseen perustuvalla järjestelmällä. Tutkimuksessa hyödynnettiin rengasäänistä koostuvaa tietoaineistoa, joka koottiin kahdelta mittauspaikalta tallennetuista äänitteistä. Rengasäänien kerääminen mittauspisteillä toteutettiin tiepäällysteen alle asennettujen kontaktimikrofonien avulla. Väyläviraston ylläpitämiä automaattisen liikenteenseurantajärjestelmän mittauspaikkoja hyödynnettiin äänitteiden keräämisessä sekä tietoaineiston koonnissa.

Renkaidentunnistusjärjestelmän suunnittelussa ja toteutuksessa sovellettiin digitaalista signaalinkäsittelyä ja koneoppimista. Henkilöautontunnistin kehitettiin osaksi järjestelmää mui-den ajoneuvotyyppien poissulkemiseksi tunnistuksesta sekä renkaiden paikantamiseksi luoki-teltavista äänisignaaleista. Piirteidenirrotus rengasäänistä toteutettiin mallintamalla ihmisille ominaista tapaa havaita ääniä. Rengasluokittelijasta kehitettiin kaksi vaihtoehtoista versiota ohjatun oppimisen menetelmiä hyödyntäen. Luokittelijat toteutettiin tukivektorikone- ja neuro-verkkotyyppisinä. Tietoaineiston rengasäänitteet annotoitiin ajoneuvoluokkien ja rengastyyp-pien mukaisesti nastarenkaisiin ja ei-nastarenkaisiin, joista henkilöautojen rengasäänitteet va-likoitiin luokittelijoiden koulutus- ja testauskäyttöön. Luokittelijoiden koulutus ja testaus toteu-tettiin mittauspistekohtaisesti käyttämällä ensimmäisen äänityskohteen data treenaukseen ja toisen äänityskohteen data testaukseen. Koejärjestelyjen tarkoituksena oli selvittää luokitteli-joiden suorituskykyä uudessa mittausympäristössä eliminoimalla mittapaikka- ja laitteistokoh-taiset riippuvuustekijät. Eri luokittelijatyyppien tunnistustarkkuuksia vertailtiin ja analysoitiin testiaineiston tulosten perusteella.

Tutkimuksessa selvisi, että ajoneuvojen rengastyyppien tunnistus on mahdollista toteuttaa automaattisesti testeissä käytetyillä mittausmenetelmillä. Sekä henkilöautojen tunnistuksessa, että rengastyyppien luokittelussa saavutettiin noin 95% tunnistustarkkuudet. Testattujen luokittelijatyyppien väliset erot luokittelutarkkuuksissa olivat pieniä. Tutkimuksen tulokset viittaavat siihen, että järjestelmä kykenee toimimaan ennalta tuntemattomassa ympäristössä yksittäisel-tä mittauspisteeltä kerätyn opetusaineiston perusteella. Mittauspaikkojen vähäisen lukumäärän takia järjestelmän laajamittaisesta toimivuudesta ei voida kuitenkaan päätellä mitään ilman jatkotutkimuksia. Jatkotutkimuksia varten tulisi kerätä lisää rengasääniä uusilta mittauspaikoilta, jolloin järjestelmän tunnistustarkkuutta ja yleistyvyyttä voitaisiin mahdollisesti parantaa.

Avainsanat: renkaiden luokittelu, nastarenkaat, automaattinen äänentunnistus, liikennetietojen keräys

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

PREFACE

This thesis was done as a sub-study for a research project about classification of studded tires. The project was a co-study between Audio Research Group and Research Group on Earth, Foundation and Railway Structures. The study was ordered by Finnish Transport Infrastructure Agency.

I want to give a special thanks for Toni Heittola, who acted as my supervisor throughout project in a professional and friendly way. I also want to thank Ville Liiv for the close copperation in the research project and for my first co-authoring acknowledgement in an academic release.

In Tampere, 29th April 2020

Aapo Hakala

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

DFT	Discrete Fourier Transform
FTIA	Finnish Transport Infrastructure Agency
MLP	Multilayer Perceptron
STFT	Short Time Fourier Transform
SVM	Support Vector Machine

1 INTRODUCTION

Snowy and icy roads make the use of winter tires mandatory in many countries [28]. The goal with the mandatory use of winter tires is to reduce the number of traffic accidents and to make driving possible on slippery roads. The traction of a winter tire is much better on snow compared to a regular tire because of the different designing in the tire tread pattern [5]. In addition to thread pattern design, metallic studs are often used because they provide better traction control on ice [24]. Studded tires are common in Nordic countries where the driving conditions can be challenging in terms of road slipperiness.

Figure 1.1. *Examples of studded (left panel) and non-studded tires (right panel).*

There are many downsides in the use of winter tires as well. Vehicles with studded tires have a significant impact on the rutting of asphalt pavements as they grind away the road surface nearly three times faster than non-studded tires [6]. Studded tires have also a negative influence on air quality causing health risks near highways and large cities due to the generated wear particles [12][15]. Ground road pavement generates street dust that makes the environment dirty and decreases the comfort of citizens [8]. These effects tend to be at their worst in springtime when the outdoor temperatures rise and there is no fixed layer of snow or ice on the road surfaces. The weather conditions can still change rapidly and cause roads to be slippery which prevents cautious drivers from changing their winter tires until the temperatures rise permanently above 0 $^\circ$ C. [22]

Finnish Transport Infrastructure Agency (FTIA) is responsible for developing and maintaining the state-owned road network. FTIA collects information about road traffic by using an automatic traffic monitoring system. The data collection for the traffic monitoring is done by roadside *traffic measurement stations* and there are about 500 measuring sites around the country. A single traffic measurement station registers bypassing vehicles and collects data about their bypass time, direction, lane, speed, vehicle length, time elapsed between vehicles and the vehicle class. However, no data is collected about the tires of bypassing vehicles. [25]

FTIA is interested about the use of studded tires on Finnish roads and has published a research study regarding the subject [16]. The objective in the study was to find out how people time the change of their winter tires with respect to the weather and driving conditions. The resulted estimations in the study were mostly based on data provided by car service companies that offer tire changing services. Another data collection method that was used was a hearing based roadside sample survey. In hearing based data collection the audible differences of moving vehicles were listened and categorised based on the sound of tires. The sound of studded tires is considerably noisier compared to non-studded ones because of the metallic studs. The scratching sound of studded tires on snowless pavement makes their recognition possible for humans and presumably for machines as well.

It is clear that the above mentioned data collection methods provide only broad estimations about the statistical distribution of studded tires. The main problem is that the data from car service companies does not include road-specific information about the vehicles. In other words, it is impossible to name the exact roads or sections of a roads where such predictions are valid. There is also no information about the numbers of private tire changing which is a common practice in sparsely populated areas [16]. Hearing based roadside sample surveys are location-specific but still inconvenient as manual work requires a lot of time and manpower in order to get large amounts of data.

A continuous and automatic data collection about the use of studded tires would be possible by installing a tire classification system as part of a traffic measurement station. Accurate measurements of a single measuring site could be used to estimate the tire distribution in the local area. A network of traffic measurement stations performing tire classification would enable nationwide monitoring of studded tires which would be a significant improvement to the existing data collection methods.

In this thesis, a prototype of an automatic tire classification system is designed and implemented by combining digital signal processing and supervised machine learning techniques. The purpose of the thesis is to present the implemented system that can be used in a traffic measurement station to collect data about the use of studded tires. The thesis is built round a comparison between two different classifier architectures, one based on *support vector machine* (SVM) and the other on *multilayer perceptron* (MLP). A theoretical introduction for the classifiers is given and the thesis goes through the implementation and the system experiments that were carried out according to supervised machine learning methods. An audio dataset of in-road recordings is exploited in the training and evaluation of the system. The thesis was done as a sub-study for FTIA's research project about classification of studded tires and the implemented system is the same than the classifier referred in the online publication of the study [14].

The structure of the thesis is the following. A theoretical overview about the machine learning methods and the audio processing techniques are discussed in Chapter 2. The second chapter also covers background information for the data by presenting the recording setup and the equipment that was used in the data collection. A detailed description about the pipeline of the implemented system is presented in Chapter 3. In Chapter 4 the content of the audio data is examined and the experiments for the implemented classifiers are carried out. The results of the experiments are analyzed in Chapter 5 and in the end the final conclusions about the experiments are presented in Chapter 6.

2 METHOD

The second chapter covers the theoretical backgrounds of the tire-classifier system. The recording setup and the equipment used in the data collection are first presented in Section 2.1. Machine learning is then introduced and discussed in Section 2.2. The machine learning section aims to provide both general level introduction and detailed explanations about the operating principles of SVM and MLP classifiers. Section 2.3 discusses about the role of features in a machine learning system and analyses the audio processing steps that are necessary in order to get meaningful input data for the classifiers.

2.1 Microphones and the recording setup

The recording of the data was done in real life environment using *contact microphones*, also known as *piezoelectric microphones*. Contact microphone is a sensor that uses *the piezoelectric effect* to measure force or strain [2]. Piezoelectric effect is a complex phenomenon that covers most of the areas of classical physics. It applies electronics, mechanics, thermodynamics, circuit theory, crystallography and elasticity and strength of materials [19]. While the underlying theory of piezoelectricity is out of the scope of this thesis, it is still worth mentioning why contact microphones are the most reasonable choice for the task at hand.

Unlike other microphone types, contact microphones have a natural ability to convert mechanical vibration of a solid substance straight to electrical signals without the need of sound waves traveling through air. By placing the contact microphones under the road pavement and close to the vehicle path the sound of tires is received much more louder compared to other sounds of the vehicle. Another benefit of in-road recording is that background noises from other vehicles are reduced along with the environmental and weather dependent sounds. The effect of acoustic reverberation and reflections are also less dominant below the road surface compared to a recording setup above the ground plane. Contact microphones area physically durable as they are made of solid crystal materials like quartz [19]. They are still relatively weak, as the maximum load for a typical piezo film is about 6-9 kg, which means it can not bear the weight of a vehicle as such [18].

Two DiMarzio DP130BK contact microphones were buried under the road pavement in each recording setup. The microphones were enclosed inside a plastic box to increase the physical robustness of the system and to isolate the microphones from external hu-

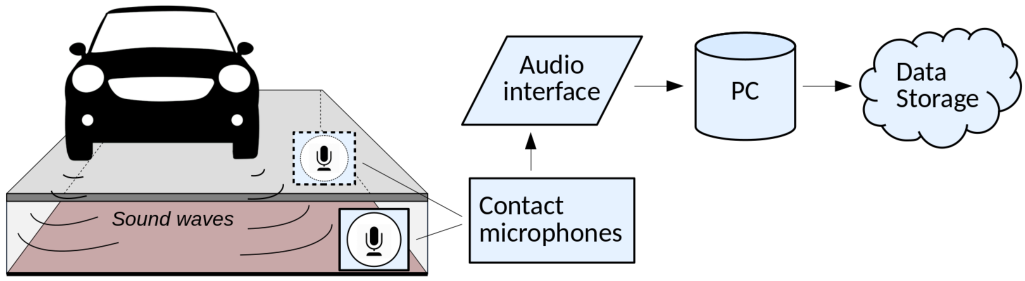

Figure 2.1. *An overview of the recording setup.*

midity and weather conditions. The microphones were then connected to a computer via Roland Rubix 44 audio interface. The audio data was recorded in wav-format using sampling rate of 44,1 kHz and bit depth of 24 bits. The recorded files were first stored to the computer memory until they were uploaded to a data storage center via 4G mobile network. An overview of the recording setup is shown in Figure 2.1.

2.2 Machine learning

Machine learning and deep learning are subcategories of artificial intelligence [29]. They provide an alternative approach for solving problems that are too complicated to be derived explicitly in a mathematical form. Machine learning and deep learning enable systems to automatically learn and improve from experience. In practice, they are data-driven computer programs that can solve complex problems such as classification and pattern recognition tasks. Machine learning is used in wide range of fields including audio related signal processing. [27][7]

There are three main types of machine learning problems: *supervised learning*, *unsupervised learning* and *reinforcement learning*. The main difference between supervised learning and the two other problems is that in supervised learning the data is provided with additional information about the desired outcome of the system. The learning is therefore supervised in the sense that the known input-output connections act as a supervisor during the learning process. The implementation of the tire-classifier is based on supervised learning. [1]

In supervised learning, the idea is to find a mapping function that best describes the connection between a given set of input-output examples. In general, supervised learning process can be expressed with a simple equation $f(X) = Y$, where X is the input data, Y is the target output for a given input and f is the desired mapping function. Such input-output examples are commonly referred as *labelled data* while the mapping function is often called by the type of the learning algorithm or architecture that is used in the system. [29]

Supervised machine learning includes several learning algorithms with different architectures. SVMs, neural networks, linear regression, decision trees, naive bayes and k-nearest neighbor algorithms are examples of supervised learning methods. The list gets even longer when all sub-categories of the methods are taken into account. Neural networks for example cover a wide variaty of implementations but the only sub-class discussed in this thesis is MLP which is a basic type of neural networ. [10]

Different learning methods have different approaches for solving problems and the resulting performances will also vary depending on the method. It is good practice to compare several options to see what method works best for a given task. In the following sections SVMs and MLPs are discussed in greater detail as they are expected to perform well in the audio classification.

In machine learning, classification is the problem of finding a good strategy to assign classes to objects based on past observations of object-class pairs. In *binary classification* there are two groups of data to which all of the objects are divided. Supervised learning is a feasible approach for solving classification tasks because of the nature of the classification problem. Object-class pairs, that is labelled data, are the known input-output values and the goal is to find out the connection between them. The connection can be modeled with a mapping function and the mapping function can be determined by using supervised learning methods.

Some supervised learning methods require large amounts of labelled data in order to learn the mapping between the inputs and the corresponding classes [7]. Labelled data is often produced manually and the annotation work can be laborious and time consuming. The data requirements are highly dependent on the learning method but a lack of labelled data usually results in a poor classification performance.

2.2.1 Support vector machine

One of the most popular algorithms in machine learning is a support vector machine. SVMs are capable of solving both classification and regression tasks. They are widely used in classification objectives as they often provide good performance on reasonably sized datasets [11][26]. Unlike many other algorithms, SVM can perform effectively with high-dimensional data because of its inner design.

The operating principle of SVM can be described as follows: given labeled training data in a feature space, SVM constructs a hyperplane as the decision boundary that separates the data samples of different classes. The samples of same class are now located on the same side of the decision boundary. The separation is done such that the margin of separation between positive and negative examples is the widest possible. The goal of the support vector machine is to find the particular hyperplane for which the margin of separation is maximized. The optimal hyperplane is determined by the training samples that are closest to the decision boundary as seen in Figure 2.2. [7]

In Figure 2.2 the data in two-dimensional feature space can be easily separated with a

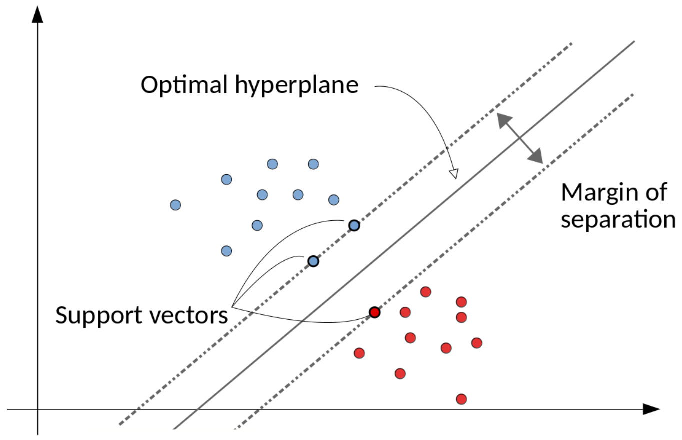

Figure 2.2. *Illustration of a support vector machine in 2D feature space. The data samples that are closest to the decision boundary are called support vectors.*

straight line. Usually this is not the case in real life as the data and the feature space can be much more complex. In order to deal with linearly non-separable data, the feature space can be transformed and modified with a method called *kernel trick*. The idea in kernel trick is to map the non-linearly separable data into a higher dimensional space where the data can be once again separated linearly. Kernel is a function that defines the inner products in the transformed space and this reduces the complexity of finding the mapping function from feature space to the output space. An illustration of kernel trick is shown in Figure 2.3. [7]

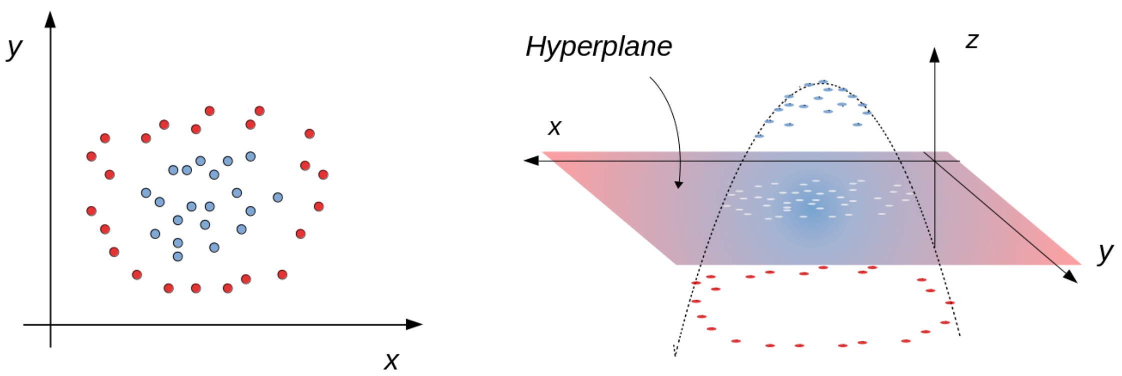

Figure 2.3. *An example of kernel trick. Linearly non-separable data in 2D feature space on the left is mapped into a higher dimensional space where linear separation is possible.*

Training of a SVM model is the process of finding the decision boundary that performs best in the classification of the labelled training data. After the training is done, the information about the kernel function and the decision boundary is stored in to the model. The model can be used later to predict the class of previously unseen and unknown data. The model transforms the given inputs into the feature space and the decision about the output class is determined by the side on which the sample lands on with respect to the hyperplane of the model. As a result, a prediction of the output class is given at the output.

SVMs have their disadvantages as well. SVMs do not work well on extremely large sets of data as they become computationally very expensive [17]. They also perform poorly with noisy data or if the target classes are overlapping. There is no probabilistic explanation for a classification result since the decision is based on to the sample position with respect to the hyperplane. [23]

2.2.2 Multilayer perceptron

Neural networks are parallel distributed processors made up of simple processing units called neurons. Neural networks have a natural propensity for storing experimental knowledge and making it available for use. The structure of the model is designed in the way in which human brain functions and performs different tasks. Like human brain, neural network acquires knowledge from its environment through a learning process. The knowledge is stored as synaptic weights in between neurons that are connected to each other. An example of a neural network is illustrated in Figure 2.4. [7]

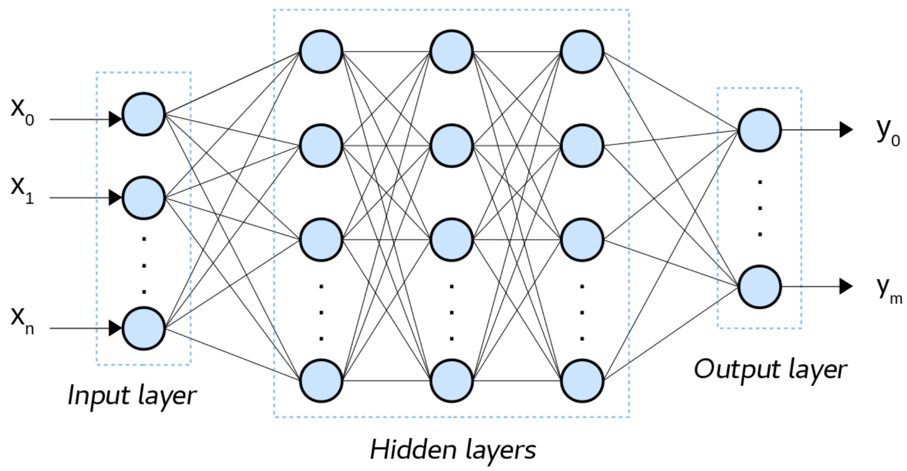

Figure 2.4. *An example of a neural network. The fully connected feedforward network has n inputs, m outputs and three hidden layers.*

In neural networks, a group of neurons that are parallel with each other form a layer. In Figure 2.4, the left-most layer of the network is called the input layer while the right-most layer is referred as the output layer. These layers pass data in and out of the network respectively. The remaining layers between the input and output layers are called hidden layers. Neural networks with one or more hidden layers are referred as multilayer perceptrons. [30]

The operating principle of a single neuron is shown in Figure 2.5. In a fully connected MLP, all neurons at the same layer are connected to every neuron at the adjacent layers.

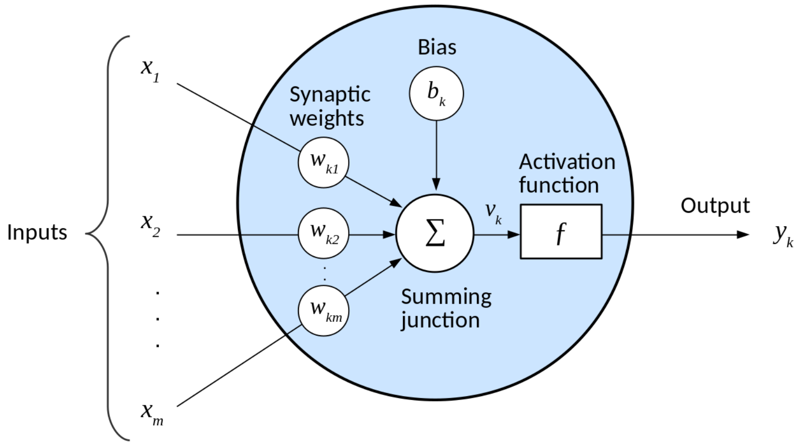

Figure 2.5. *A neuron and its functionality after S. Haykin [7].*

Therefore the inputs of a neuron are the outputs from the neurons at the previous layer. Each connection between two neurons have their own weight parameter w_k. At first, the inputs are multiplied by their weight parameter correspondingly. The resulting products are added to produce a weighted sum v_k. The weighted sum v_k includes a bias term b_k that is used to modify the importance of the neuron. Finally, an activation function is applied to the resulted sum to introduce non-linearity into the output of a neuron. [7]

Training of a neural network can be seen as an optimization problem. The objective in the training is to find values for synaptic weights and biases that provide the best performance for a given task. The optimal combination of weights and biases can be found by introducing a *cost function*, which is a measure of performance of the network. Cost function quantifies the error between expected output and measured output values. The smaller the error produced by the cost function, the better the performance of the system. There are several cost functions that can be used to measure the performance. Independent of their mathematical implementation, they take as many inputs as there are weights and biases in the network and output a single cost value. [7]

An intuitive analogy for a cost function is to think of it as a hilly surface. The cost value determines the height of the surface while the weights and biases refer to other coordinates in the space. The goal in the training is now to find the coordinates to the deepest spot in the valley of cost values. In practice the number of neurons is often huge so the number of dimensions explodes making the input space huge as well. Consequently, the cost surface is most likely to be non-convex meaning that there are multiple tops and bottoms with various heights. The optimal solution, that is the global minimum of the cost function, might never be found because of the large number of dimensions and size of the search space. However, there are local minimas that are close to the optimal solution, so the performance of the network may be good enough with a suboptimal solution.

Gradient decent is an effective learning algorithm that can be used to iteratively adjust the weights and biases so that the overall network cost decreases gradually. At the beginning of the training the weights and biases are often initialized with random values. The superiority of the cost function at a random starting point is unknown and so is the environment around it. Weight-specific instructions for tuning the parameters are determined by the negative gradient of the cost function. Negative gradient is a vector on the cost-surface that shows the direction of steepest down hill. A step into the direction of negative gradient will decrease the cost most efficiently. The gradient of the cost function can be calculated with *backpropagation* technique. By minimizing the cost function the network is forced to find such values for weights and biases that provide low cost and thus high performance for the given task. [21]

The number of neurons and the structure of layers have a crucial role in the performance of MLP classifier. A network with only few neurons does not have the capacity to store large amounts of information to its neurons. Then again a network that is too large may suffer from overfitting, in which case the network is unable to generalize its knowledge to new data and it performs only well with the data it has seen before in the training. In order to produce meaningful information, it is reasonable to match the size of the input layer with the size of the input data while the neurons at the output layer should present the desired output such as classes in case of classification. This way the network outputs can be considered as predicted probabilities for the corresponding classes.

2.3 Features

The performance of a supervised machine learning system is highly dependent on the data. This applies particularly to classification problems as the input data contains all the information that is needed for the solution. In theory, if a classifier is trained with extremely many training samples it eventually learns to identify the given classes without errors. In practice the amount of data is always limited so it is desirable to speed up the learning process if possible. This can be done by presenting the data so that the essential properties of the classes are emphasized.

In the context of machine learning, *a feature* is a measurable property of an object that is being analysed. Features are obtained from the input data in a process of *feature extraction*. In feature extraction the data is transformed into a feature space. This transformation may be viewed as a dimension reduction or data compression that simplifies the task of classification. Feature extraction can be done in multiple ways and it is usually application specific. The essential properties of the data should be taken into account when determining the features and the feature extraction procedure. [7]

Spectral properties are often examined in the feature extraction stage of audio related machine learning systems. Spectral properties of an audio signal determine the nature of the sound and the way the sound is perceived. Unique sounds have a unique spectrum while similar sounding signals have similar spectral properties. Therefore the spectral components of sounds can be considered to be essential features in audio data.

There are also inessential properties in audio that may affect negatively to the performance of the system. In audio classification such inessential properties may occur from background noise and other external factors like reverberation and phase differences. Inessential properties should be also taken into account when designing the feature extraction steps by minimizing their effect if possible.

When extracting spectral features from an audio signal, a frequency-domain presentation of the signal is required. Audio signals are typically presented in time-domain where the data is shown as signal magnitudes with respect to time. In frequency-domain presentation the time-axis is replaced with a frequency-axis and the signal is shown as frequency components from which the signal consists of. The process of decomposing a signal or a function to its constituents is called *Fourier transform*. In order to transform discontinuous data, *discrete Fourier transform* (DFT) is needed due to the discrete nature of digital signals. Hovever, DFT alone is often inconvenient for non-stationary signals such as audio data. It produces averaged frequency information over the entire time interval of a given signal. Therefore the time-specific information is lost as the time-dimension no longer exists in the frequency-domain.

Short-time Fourier transform (STFT) is a common method to transform non-stationary signals from time-domain into frequency-domain. STFT describes the evolution of frequencies of a signal with respect to time. In STFT the signal in time-domain is first split into a sequence of overlapping and equally sized blocks of data. Each block is then multiplied by a windowing function to minimize spectral leakage in the following domain transform. After applying DFT to each windowed block, the resulted frequency-domain vectors are concatenated to form the STFT-matrix. Each column in the matrix contains the frequency information of the signal at a specific time interval. Each row is a narrow frequency band describing the evolution of the signal components within the range of the band. [13] An example of STFT presentation is shown in Figure 2.6

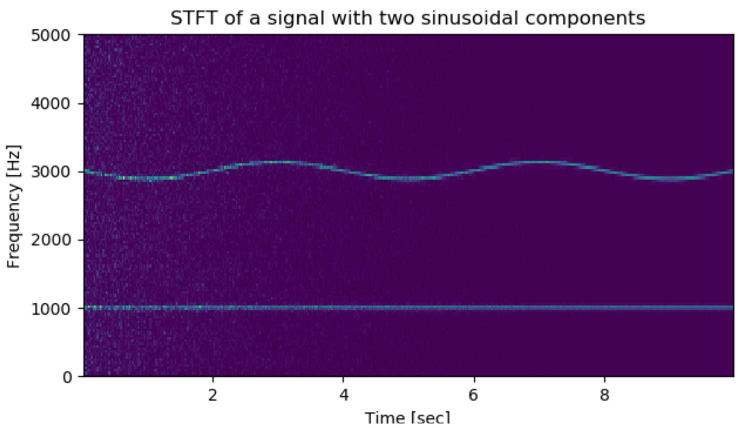

Figure 2.6. *Example of STFT presentation. The signal contains two sinusoidal components with 1000 Hz and 3000 Hz frequencies. The frequency of the component with the higher pitch is slightly alternating which causes it to fluctuate in the graph.*

Audio signals are often presented visually with a *spectrogram*, which can be obtained by computing the squared magnitudes of the STFT of a signal. By analyzing the spectrogram it is possible to identify the characteristics of the signal and the soundscape of the recording in general. STFT-matrix as well as the spectrogram presentation provide time-localized frequency information which makes them both useful for feature extraction purposes.

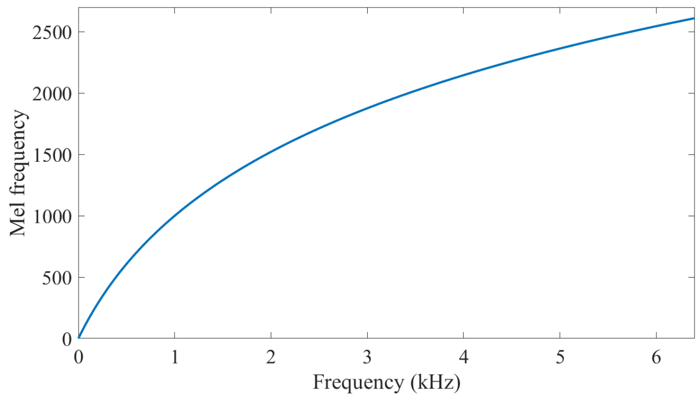

Figure 2.7. *Illustration of mel frequency scale with respect to linear frequency scale after Aalto University Wiki [4]*

Feature extraction from audio can be designed to model human hearing. *Log mel-spectrogram* is a type of spectrogram that has a logarithmic amplitude axis and a *mel scaled* frequency axis. Mel-scale approximates the way in which humans perceive frequencies along the audible frequency range. Illustration of the mel-scale is shown in Figure 2.7. Human auditory system is more sensitive to lower frequencies and this is taken into account by presenting the spectrogram in a mel-scale. The logarithmic magnitude scale also models the human auditory system as the perceived loudness of a sound behaves in a logarithmic manner [9].

3 SYSTEM DESCRIPTION

In this chapter the implementation of the system is described. The system is divided in three parts that are *Passenger car detector*, *Feature extractor* and *Classifier*. The operating principle of each part is presented in the corresponding sections 3.1, 3.2 and 3.3. The practical implementation of the system has been coded in python programming language. The pipeline of the system is shown in Figure 3.1.

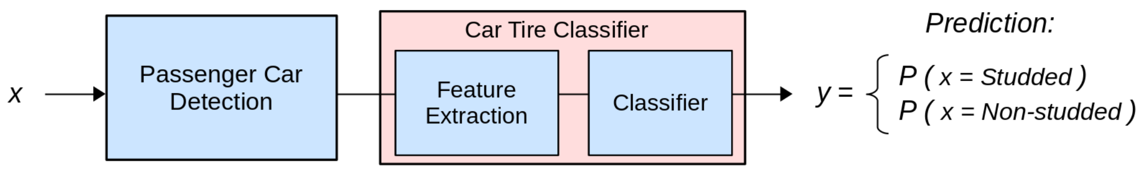

Figure 3.1. *The pipeline of the system.*

3.1 Passenger car detector

In order to do the binary classification for passenger cars, the system must be able to detect and separate them from other traffic. The passenger car detector is responsible for feeding the classifier with right kind of data. Whenever an input audio file contains the sound of a passenger car, the detector should pass the data for the feature extraction stage. A block diagram of the passenger car detector is shown in Figure 3.2.

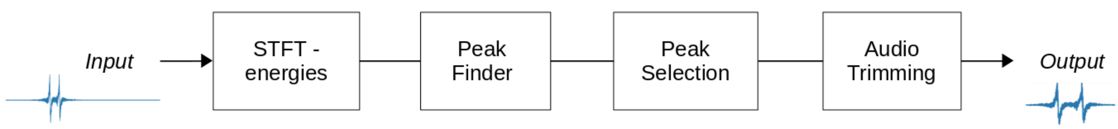

Figure 3.2. *Block diagram of the passenger car detector.*

The passenger car detector gets about four seconds long audio recording as its input. The signal is first normalized and the STFT of the normalized signal is computed. For each time-frame in the STFT-matrix the mean of squared frequency-bins is calculated. Consequently, all frequencies within each time-frame are merged together into a series of scalar values. By concatenating all of the scalar values into a vector, averaged STFT energies are obtained. The resulted vector represents the spectral energies of the signal with respect to time. A visualization of the procedure is shown in Figure 3.3.

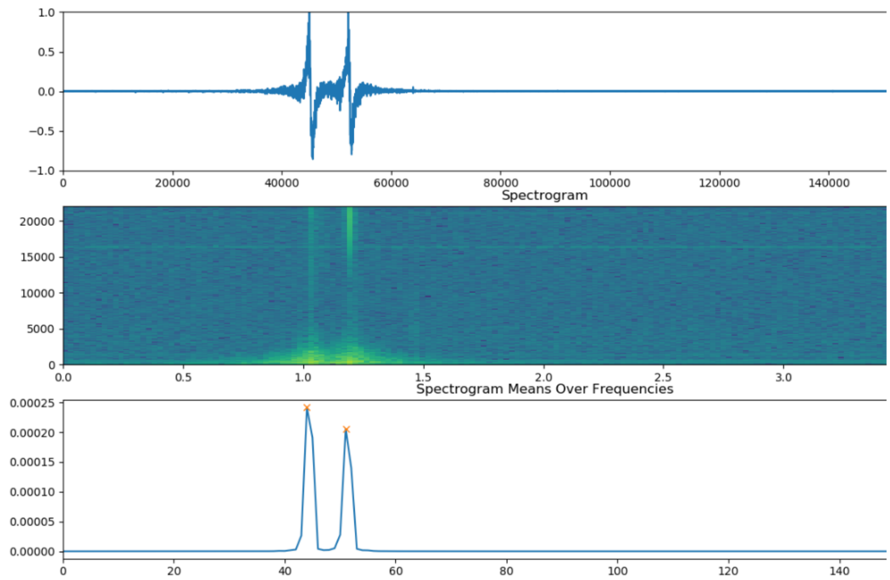

Figure 3.3. *Waveform of a recorded signal presented in time-domain, frequency-domain and squared averages.*

A peak finder-algorithm is then utilized to spot the high-energetic frames from the averaged STFT vector [20]. The peak finder detects the local maximas from the input sequence as visualized in the lowest panel of Figure 3.3. A list of time-indexes whose value exceeds a certain threshold is produced to its output. An audio recording that contains one passenger car should now generate two time-frames, one for the front tires and the other for the rear tires.

The peak finder-algorithm makes the passenger car detector sensitive for impulsive noises so the number of output elements produced by the peak finder is often greater than the number of axles of vehicles in the recording. A peak selection-block is used to identify the real axles of the recorded vehicle from possible false-detections. Most common reasons for a false-detection are vehicles on the opposite lane and digital noise created by the data transmission.

The peak selection-block consists of lane inspector, digital noise-filter, unpaired peak-detection and an examination about double detection of the same axle. Each sub-task is designed to detect and remove the specific false-detection type. Lane inspector detects the vehicles on the opposite lane by comparing the amplitudes of the detected peaks and removing a peak that has less energy compared to others. Digital noise-filter searches for peak-frames that have a sudden pulse-like signal waveform. In unpaired peak-detection the parity of the detected peaks is checked and if the number of peak-frames is odd, the peak that stands out from the other two peaks is removed. Double detection of a same axle can be spotted from the short distance between the detected peaks.

The final conclusion about the presence of a passenger-car is based on the output of the peak selection-block. In a positive detection case the number of remaining peaks after the peak selection is equal to two and the time difference between the two peaks is less than 0.4 seconds. The peaks determine the exact bypass moment of the axles so larger vehicles can be ruled out by setting a limit for the time difference of the peaks. In a negative detection, the number of peaks is other than two or the time difference between two detected axles exceeds the 0.4 seconds limit.

A signal that generates a positive detection is sliced into a short segment of audio. The idea is to remove the irrelevant parts of the recording and preserve only the sound of the detected tires. The cutting takes place in the audio trimming block which gets the time stamps of the tires from the peak selection block. The time stamps are transformed to the corresponding sample index values in the original time-domain signal and a trimmed version of the input audio is produced. The output of the audio trimmer block is also the final output of the passenger car detector.

3.2 Feature extractor

Features are extracted from the trimmed audio data. Log mel-spectrogram of a signal consists of 40 frequency bins in the frequency axis. For each row in the log mel-spectrogram, the mean of the amplitudes along the time-axis and the standard deviation of the mean are calculated. The resulting means and standard deviations are gathered into a feature vector. The feature vector contains 40 means and 40 standard deviations and the value of each element is based on the content of the log mel-spectrogram of the signal. After the feature extraction the feature vector is sent to the classifier for the classification. An example of a log mel-spectrogram is shown in Figure 3.4.

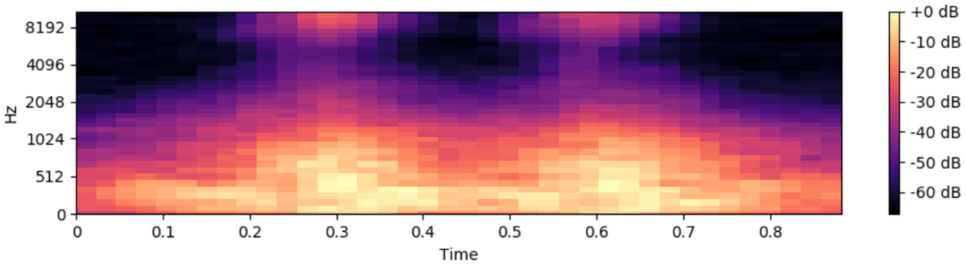

Figure 3.4. *Log mel-spectrogram of an audio signal that contains a vehicle with studded tires. The effect of studs can be seen in the upper most frequency bins.*

3.3 Classifier

The classification into studded and non-studded classes is done in the classifier-block. Two individual implementations of the binary-classifier were trained, one based on SVM and one using MLP. Both models use the same feature-vectors as their inputs and both models output a prediction about the class of the input data. In addition to the output class, the models also produce a likelihood value that indicates the probability of the prediction. The higher the likelihood value the more certain the model is about the class of the input data.

The SVM-model uses a linear kernel-function. A regularization parameter, which is a measure of importance of mis-classifications during the training, was set to 0.0001. The likelihood value of a predicted output class is based on the shortest distance from the sample in the feature space to the decision boundary of the model. The likelihood value is not equivalent with the real probability of the prediction but it is useful in detecting the borderline cases.

The MLP-model is a fully connected feedforward network with six hidden layers. The number of neurons at each layer including the input and output layers are 80, 42, 41, 31, 22, 13, 6 and 1. Adam-optimizer, which is an adaptive gradient descent optimization algorithm, was used in the training of the MLP-model. Rectified linear unit-activation function is used at hidden layers while the output layer uses logistic sigmoid-function. The output of the MLP-model presents also the likelihood value of the predicted class with a threshold of 0.5.

The system runs all of the operations channel-wise meaning that the data from each microphone is processed separately. A single vehicle may produce one or two classification results depending on the performance of the passenger car detection. If the channel-wise results for the same vehicle are inconsistent, the final decision about the output class is based on maximum likelihood. The class-specific likelihood values are compared and the output of the channel with the highest likelihood value is chosen as the output of the classifier.

4 EXPERIMENTS

The fourth chapter covers the implementation of the experiments that were carried out and illustrates the final results of the experiments. The data collection process and the recorded audio content are first analysed in Section 4.1. Section 4.2 explains the division of the data into training set and test set and shows the statistics about the data used in the experiments. In Section 4.3 the results of the experiments are presented in a table format.

4.1 Data

Figure 4.1. *Recording locations on map.*

The data used in the experiment was recorded in two different locations in Saaramaa and Alavus by the major roads VT26 and KT66 (Figure 4.1). The data was collected on several days during the tire transition seasons in autumn 2018 and spring 2019. A typical tire transition season in Finland lasts for over a month and in spring time even longer. The data was collected sparsely along the seasons in order to get a sufficient number of recordings from both tire classes.

The recorded data was listened through and annotated manually by a single person. In the annotation process the audio files were first labelled based on the type of the vehicle. The recordings containing passenger cars were then classified into studded and non-studded classes. The labeling of the two classes was done in a stepped manner to indicate how obvious or ambiguous each sample was to classify. Each file was rated with an integer value from a sliding scale of 1-5. Values 1 and 5 stand for clear cases of 'studded' and 'non-studded' classes respectively. The samples labeled with a value of 2 are likely to belong to 'studded' class, while the samples labeled as 4 are likely to go to 'non-studded' class. The rating value 3 means that the audio file is very ambiguous and difficult to classify for a human. These files were discarded from the final dataset to minimize the number of outliers in the experiments.

Over 4200 recordings of annotated passenger cars were chosen to the dataset for the experiments. Almost half of the data from the original collection had to be excluded, such as recordings containing multiple cars that proved to be difficult to annotate. Other vehicle types as well as recordings with extremely noisy data were also left out from the dataset for the experiments.

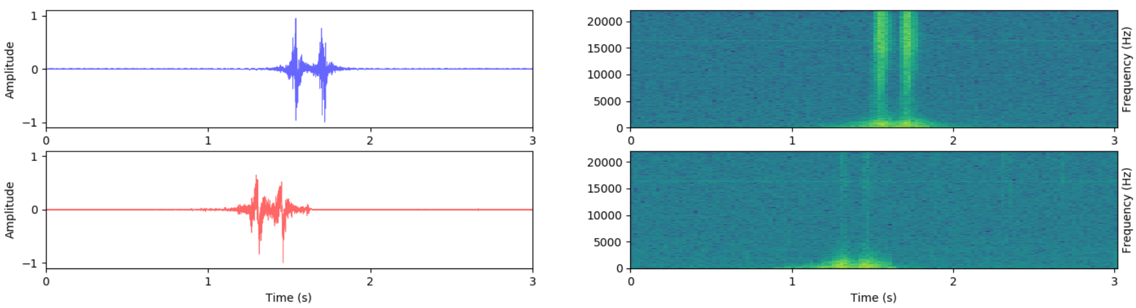

Figure 4.2. *Waveforms and spectrograms of vehicles with studded (upper) and non-studded (lower) tires. The spectrograms of the signals show that studded tires have more energy on high frequencies compared to non-studded tires.*

Examples of recorded vehicles are shown in Figure 4.2. In both recordings the vehicle is passing by the microphone at about 1.5 seconds. The vehicles are seen as two high-energetic amplitude peaks that are created by the tires of the vehicle. As a result of the recording geometry, the sounds of tires that are attached to the same axle of the vehicle are merged together into a single amplitude peak. Thus the first impulse on the left corresponds to the sound coming from the front tires while the latter amplitude peak is generated by the rear tires.

Both tire classes are shown on top of each other in Figure 4.2. It is hard to tell from the time-domain presentation which signal is created by studded tires and which one belongs to the non-studded class. There are still audible differences between the recordings that can can be visualized in the frequency-domain. The spectrograms of the signals are shown next to the corresponding waveforms. By analysing the spectral properties of the signals it is easier to tell the difference between the two classes.

Recorded audio files are about four seconds long in duration. The automated recording process records multiple seconds of background noise in the beginning and at the end of each file. The speed of the recorded vehicle determines the by-pass time so the timeline is never fixed. The classifier needs to focus on the exact moment of by-pass and ignore the remaining parts of the recording. The more precise the time window around the vehicle, the less irrelevant information is fed to the classifier.

4.2 Evaluation

The data for the experiences set was split into a training set and a test set. The statistics about the data is shown in Table 4.1. The division into training set and test set was based on the location of the recordings. The training set contains only data from KT66 while the test set consists of data from VT26. This way the generalization of the system can be verified as location dependent factors in the recordings are different between the two sets. The amount of data and the fact that the quality of the data from KT66 is better also affected to the decision about the train/test-split. The system is expected to learn the essential properties better from the data that has less noise in it.

	Training (KT66)	Testing (VT26)	Total
Studded	1639	717	2356
Non-studded	1214	638	1852
Total	2853	1355	4208

Table 4.1. *Number of annotated recordings from each location. The data from KT66 is used in the training and data from VT26 as testing set*

4.3 Results

The performance of the implemented systems was evaluated with the test set (VT26 data). Passenger car detector was first applied to the test set. After passenger car detection the, files that triggered a positive detection were fed to the classifier-models. The files from which no passenger car were detected were discarded. The results of the experiments are shown in the corresponding subsections 4.3.1 and 4.3.2.

4.3.1 Passenger car detection

The detection results of the passenger car detector are shown in Table 4.2. The passenger car detector was able to recognize 1300 vehicles out of the 1355 passenger cars in the test set. Out of the 55 passenger cars that triggered a negative detection, 27 were using studded tires and 28 non-studded tires.

	Training (KT66)	Testing (VT26)	Total
Studded	1627	690	2317
Non-studded	1194	610	1804
Total	2821	1300	4121

Table 4.2. *Number of detected passenger cars from each location.*

	Training (KT66)	Testing (VT26)	Total
Studded	0.993	0.962	0.983
Non-studded	0.984	0.956	0.974
Total	0.989	0.959	0.979

Table 4.3. *Class and dataset-specific detection accuracy of the passenger car detection.*

4.3.2 Car tire classification

The results of the SVM and MLP classifiers are shown in Table 4.4. Both of the classifier reached over 94% classification accuracy. The number of true studded and true non-studded predictions are shown in the corresponding rows to present the class-specific prediction accuracy of the models. The false predictions of the models are also shown to present the numbers of mistakes by the models.

N=1300	SVM	MLP
Accuracy	**0.965**	0.946
True studded	679	570
True non-studded	575	590
False studded	35	20
False non-studded	11	120

Table 4.4. *Classification results of the implemented classifiers. N is the number of vehicles in the test set after the passenger car-detection. True predictions show the number of correct predictions while false predictions indicate the class that was incorrectly predicted by the classifier.*

5 DISCUSSION

In this chapter the results of the experiences are discussed. A comparison between the implemented classifier-models is made by analysing their performance on the test set. Some of the challenges that were faced during the implementation of the system are presented along with possible improvements for further development. A discussion about the performance of the passenger car detector is covered in Section 5.1 and Section 5.2 compares and analyses the results of the classifiers.

5.1 Passenger car detection

The passenger car detector is able to detect 95.9% of the vehicles from the test set. The detection rate was obtained by trial and error as data from both locations was used in the designing of the detector. The resulted performance is arguably good since only dozens of vehicles are missed. However, because data from both locations was used in the development of the passenger car detector, it is only efficient in the exact recording locations until proven otherwise.

The early versions of the passenger car detector were developed by using only data from the training set. The primary goal was to detect passenger cars effectively from the training set and see if the same detector would also apply to the test set. Once the data from the other location was tested it was clear that the two locations were very different in terms of signal quality. The peak finder-algorithm was too sensitive to the noise in the test set which caused the detection rate to be poor.

A choice had to be made between a truly generalizing passenger car detector and the amount of evaluation data in the classification. The high demand of labelled data in supervised learning methods made it necessary to prefer an effective detection mechanism over the experiment of generalization. Since the early implementations had failed, it was reasonable to exploit both datasets in the development of the passenger car detector. The flaws of the detector were fixed successfully by adding the peak selection unit which increased the performance significantly.

The detection of the passenger cars could also be implemented by using supervised machine learning methods. A simple sound event detector combined with a vehicle-classifier might even perform better than the existing detector. A traditional signal processing-based approach was preferred for the passenger car detection mainly because the task was expected to be easily practicable. A machine learning based detector would also require additional annotation work which is why the more straightforward method was chosen. If the implemented passenger car detector is ever replaced with a machine learning-based system, the current detector can be used as an annotation tool to speed up the process of creating labelled data.

5.2 Classifier comparison

Both of the classifier models perform very well in the tire classification task. The SVM-model classifies the test set with 96.5% accuracy while the MLP-model predicts 94.6% of the data correctly. It is clear that both classifiers are able to generalize from the training data and learn the difference between vehicles with studded and non-studded tires.

When analysing the performance of a supervised learning system, a typical question to be asked before and after the experiments is that is there enough data for the training and evaluation of the system. Apparently the amount of data is suitable for both models as seen from the good results. There is still some room for improvement and one possible way to improve the performance would be to increase the size of the dataset by adding more labelled data to it.

If the size of the dataset was increased the MLP-model would probably benefit more from that compared to the SVM-model. The SVM-model could even suffer from a performance loss due to the increased complexity of its decision boundary. With a larger training dataset it would be possible to add more hidden layers to the MLP-model without having to worry about possible overfitting during the training. The network would then be able to extract more detailed information about the audio features which again might improve the overall classification accuracy. Even if the network structure was kept the same the MLP-model would benefit from an increased dataset as it would gain more experience from vehicles that are tricky to classify with the current network.

The classification results in Table 4.4 show a detailed comparison about the performance of the two models. A key observation about the results is that the SVM-model is much better at recognizing studded tires. The SVM-model makes only 11 false predictions with samples of studded-class while the corresponding number of errors made by the MLP-model is 120 false predictions. This is where the SVM-model outperforms the MLP-model decisively. At the same time the statistics about non-studded class are in slight favor of the MLP-model. The margin with non-studded class is considerably smaller and not enough to compensate MLPs inaccuracy with studded-class.

Based on the results, the MLP-model seems to be slightly biased towards the non-studded class causing it to make a more false predictions with one class while performing well with the other. Possible explanations for the phenomenon are that the MLP-model is more sensitive to the location dependent factors or that the differences between the data from different locations are seen as features of non-studded class. More specifically, some neurons in the MLP-model may emphasize or de-emphasize certain features that affect directly to the classification outcome. As a result of that, the model will favor the non-studded class and predict it more often in close situations. However, because of the size of the test set, it is impossible to draw such conclusions.

One possible solution for the above-mentioned problem would be to form the training set and test set such that data from both recording locations is present throughout the whole learning process. This way the classifiers would learn the essential features of the tire-classes from a wider perspective as they would have to take into account the different circumstances of the two locations. There would also be a smaller number of confusing data samples in the test set since the classifiers would have had a chance to gain experience from similar examples. Therefore it is quite likely that the classification results would improve if the data from the two locations was split evenly between the training set and the test set. On the other hand, if the tire classification system is put into operation, such scenario is unlikely that the model would be explicitly trained with the data from the same location. In order to successfully productize the system, the classifier model needs to be able to adapt to an unknown environment.

More advanced machine learning methods could be also used instead of the implemented SVM- and MLP-classifiers. For example, *convolutional neural networks* and *recurrent neural networks* are examples of considerable alternatives for the task. Convolutional neural networks are good at recognizing patterns from two-dimensional data so they could use the entire log mel-spectrogram of the audio signals in the classification instead of handcrafted feature vectors. Recurrent neural networks are designed to extract sequential information from the data which makes them useful in audio related applications. In addition to alternative supervised learning methods, a combination of different methods, such as convolutional recurrent neural networks, are another possibility to be considered in the future development [3].

6 CONCLUSIONS

A prototype of an automatic tire-classifier system was designed, implemented and tested with promising results. The system is able to detect passenger cars from traffic flow and determine whether the vehicle is using studded or non-studded tires. The operation of the tire classification system is based on in-road audio recordings of moving vehicles on a paved road. An audio dataset of passenger cars was exploited in the training and evaluation of SVM and MLP binary classifiers. Experiments were carried out by using the dataset. A comparison between the classifier models was made based on the results of the experiments. The performances of the system implementations were analysed and ideas for possible future development were presented.

An automatic tire-classifier system was designed, implemented and tested with promising results. The system is able to detect passenger cars from traffic flow and determine whether the vehicle is using studded or non-studded tires. The operation of the tire classification system is based on in-road audio recordings of moving vehicles on a paved road. An audio dataset of passenger cars was exploited in training and evaluation of SVM and MLP classifier models. The SVM-model outperformed the MLP-model by scoring 96.5% test accuracy while the MLP-model reached 94.6% classification accuracy.

One of the main goals for the experiments was to see if the classifiers were able to adapt their knowledge to similar data that was recorded in a different location. The results of the experiments prove that both of the classifiers succeeded in the task of detecting and classifying previously unseen data in an unknown environment. The results also prove that the prototype is able to compete with an experienced human in the classification task despite the noisiness of the test data. The implemented system shows that automatic tire classification is possible and that the practical realization of such system can be done with reasonable efforts.

It is likely that the performance of the system can be improved with further research. One potential way to improve the performance is to collect and annotate more data either from the existing measurement sites or by installing the recording setup and the tire classification system to new locations. Experiments with data form new locations would help in the development of the classifier towards a well generalizing system. Another option is to redesign the passenger car detector and the classifier-model with more advanced supervised machine learning methods. However, the alternative methods require much more data so it would be beneficial to increase the number of measurement sites in order to maximize the benefits of follow-up research.

The implemented system can be installed as part of a roadside traffic measurement station or used as a stand-alone system to continuously collect data about the use of studded tires. The prototype as such is an improvement to the previous data collection methods that are irregular and broad compared to the proposed method. By installing tire classification systems into several locations it would be possible to make accurate predictions about the distribution of tire types both locally and nationwide. In the long run, the lifespans of road pavements could be estimated more accurately by monitoring the traffic and the use of studded tires on a particular road. The formation and the amount of street dust could be predicted in urban areas which would prevent unnecessary health risks. Better awareness and knowledge about the seasonal use of winter tires could help in planning of large infrastructure projects and provide advice for allocating resources for maintenance and cleaning of roads.

REFERENCES

[1] J. Brownlee, ed. *14 Different Types of Learning in Machine Learning*. 2019. URL: https://machinelearningmastery.com/types-of-learning-in-machine-learning/.

[2] A. Butterfield and J. Szymanski. *A Dictionary of Electronics and Electrical Engineering*. Oxford University Press, 2018. ISBN: 9780198725725. URL: https://www.oxfordreference.com/view/10.1093/acref/9780198725725.001.0001/acref-9780198725725.

[3] E. Çakır and T. Virtanen. *End-to-End Polyphonic Sound Event Detection Using Convolutional Recurrent Neural Networks with Learned Time-Frequency Representation Input*. 2018. URL: https://libproxy.tuni.fi/login?url=https://search-proquest-com.libproxy.tuni.fi/docview/2072286880?accountid=14242.

[4] *Cepstrum and MFCC*. Aalto University. 2019. URL: https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC.

[5] *Chassis Handbook Fundamentals, Driving Dynamics, Components, Mechatronics, Perspectives*. 1st ed. ATZ/MTZ-Fachbuch. Vieweg+Teubner Verlag. ISBN: 3-8348-9789-2.

[6] N. Fukuzaki, T. Yanaka and Y. Urushiyama. Effects of studded tires on roadside airborne dust pollution in Niigata, Japan. *Atmospheric Environment (1967)* 20.2 (1986), 377–386. ISSN: 0004-6981. DOI: https://doi.org/10.1016/0004-6981(86)90041-7. URL: http://www.sciencedirect.com/science/article/pii/0004698186900417.

[7] S. Haykin. *Neural Networks and Learning Machines, 3/E*. Pearson Education India, 2010.

[8] *Health effects of air pollutants*. Helsingin seudun ympäristöpalvelut. 2015. URL: https://www.hsy.fi/en/residents/theairyoubreathe/information-on-breathing-air/Pages/health-effects-of-air-pollutants.aspx.

[9] Hearing - from sensory processing to perception. International Symposium on Hearing Corporate Author. Springer. ISBN: 1-281-35296-9.

[10] M. Heller, ed. *Supervised learning explained*. 2019. URL: https://www.infoworld.com/article/3403403/supervised-learning-explained.html.

[11] M.-W. Huang, C.-W. Chen, W.-C. Lin, S.-W. Ke and C.-F. Tsai. SVM and SVM Ensembles in Breast Cancer Prediction. *PLoS ONE* 12 (2017). ISSN: 1932-6203. URL: https://doaj.org/article/a53c1cc0fff9462582da9d104ce3b95d.

[12] H. Karlsson, J. Lindbom, B. Ghafouri, M. Lindahl, C. Tagesson, M. Gustafsson and A. G. Ljungman. Wear particles from studded tires and granite pavement

induce pro-inflammatory alterations in human monocyte-derived macrophages: a proteomic study. *Chemical research in toxicology* 24.1 (2011). ISSN: 0893-228X.

[13] N. Kehtarnavaz. Digital Signal Processing System Design (Second Edition). Ed. by N. Kehtarnavaz. 2008. ISBN: 978-0-12-374490-6. URL: http://www.sciencedirect.com/science/article/pii/B9780123744906000076.

[14] V. Liiv, A. Hakala and P. Kolisoja. *Nastarengasluokitin*. Väylävirasto, 2020. URL: https://julkaisut.vayla.fi/pdf12/vj_2020-19_nastarengasluokitin_web.pdf.

[15] J. Lindbom, M. Gustafsson, G. Blomqvist, A. Dahl, A. Gudmundsson, E. Swietlicki and A. G. Ljungman. Wear Particles Generated from Studded Tires and Pavement Induces Inflammatory Reactions in Mouse Macrophage Cells. *Chemical Research in Toxicology* 20.6 (2007). PMID: 17516662, 937–946. DOI: 10.1021/tx700018z. URL: https://doi.org/10.1021/tx700018z.

[16] M. Malmivuo. *Talvirenkaiden käytön aikataulu ja siihen vaikuttavat tekijät*. Liikennevirasto, 2017. URL: https://www.doria.fi/bitstream/handle/10024/143921/lts_2017-40_978-952-317-445-0.pdf?sequence=2&isAllowed=y.

[17] S. Marsland. *Machine learning : an algorithmic perspective*. Second edition. Chapman & Hall/CRC machine learning & pattern recognition series. Boca Raton, FL: CRC Press is an imprint of Taylor & Francis Group, 2015. ISBN: 9781466583337.

[18] *Piezo Film Sensors - Technical Manual*. 2006. URL: https://www.contactmicrophones.com/techman.pdf.

[19] *Piezoelectric Transducers and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, 1–532. ISBN: 9783540775072. URL: https://link.springer.com/content/pdf/10.1007%2F978-3-540-77508-9.pdf.

[20] *scipy.signal.find*$_{peaks}$. Scipy. 2019. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html.

[21] A. Sharma. Guided Stochastic Gradient Descent Algorithm for inconsistent datasets. *Applied Soft Computing* (2018). ISSN: 1568-4946. URL: http://www.sciencedirect.com/science/article/pii/S156849461830557X.

[22] *Street Dust*. Finnish Meteorological Institute. 2019. URL: https://en.ilmatieteenlaitos.fi/street-dust.

[23] D. Subramanian, ed. *Support Vector Machine (SVM): A Simple Visual Explanation — Part 1*. 2019. URL: https://medium.com/towards-artificial-intelligence/support-vector-machine-svm-a-visual-simple-explanation-part-1-a7efa96444f2.

[24] *Tires for winter driving*. Tire and Rubber Association of Canada. 2019. URL: https://www.tracanada.ca/wintertirevideos/whatyoushouldknow.html.

[25] *Traffic measurement station*. The Finnish Transport Agency. 2017. URL: https://vayla.fi/web/en/open-data/materials/tms-data#.XbgMatVS-Uk.

[26] P. Wei, F. He, L. Li and J. Li. Research on sound classification based on SVM. *Neural Computing and Applications* (2019). ISSN: 09410643.

[27] *What is Machine Learning? A definition*. Expert System. 2017. URL: https://expertsystem.com/machine-learning-definition/.

[28] *Winter Tires Within Europe, In Iceland and Norway*. European Consumer Centre France. 2014. URL: https://www.ecc.fi/globalassets/ecc/ajankohtaista/julkaisut/winter_tires_2014.pdf.

[29] J. Wu, ed. *AI, Machine Learning, Deep Learning Explained Simply*. 2019. URL: https://towardsdatascience.com/ai-machine-learning-deep-learning-explained-simply-7b553da5b960.

[30] Q. J. Zhang. *Neural networks for RF and microwave design*. Artech House microwave library. Norwood, Mass: Artech House, 2000. ISBN: 1-58053-100-8.