Applying Artificial Intelligence on an Integrated Pressure and Mass Flow Sensor

Applying supervised machine learning algorithms on the data of the integrated pressure and mass flow sensor to classify the fluids inside the sensor

By

R.R.A. Groen

Supervisors

dr.ir. D. Alveringh
dr. V.D. Le
prof.dr.ir. J.C. Lötters

Bachelor thesis for Electrical Engineering

Integrated Devices and systems
University of Twente
04-06-2020

Contents

1 Abstract

The integrated mass flow and pressure sensor consist of four read-out structures, two capacitive sensors for the displacement of the tube and two resistive pressure sensors to attain the pressure drop across the tube. The sensor's purpose is to measure the mass flow in the tube and the pressure drop across the tube. However, to attain these quantities, information is lost which is related to the physical quantities of the fluid. For example, the actuation frequency is filtered but is related to the density of the fluid. This research uses machine learning algorithms to learn classification models for the fluids based on the lost information.

This is the first research into combining artificial intelligence into the field of the mass flow sensor. The first step is classifying fluids using the data. However, integrating parameter and composition estimations could be included in future research. The integration holds vast potential such as creating applications for the medical and industrial market. Applications include oil quality estimations and drug administration. Integrating artificial intelligence could potentially enhance these applications and lead to novel applications.

The training data consisted out of the non-filtered data attained from the sensor using six fluids, eight mass flows and three pre-pressures. Three machine learning algorithms are tested with the best performance achieved by k nearest neighbour and decision tree algorithms which were able to classify fluids with an accuracy of 95% and 92% respectively.

However, this accuracy is only achieved for discrete mass flows, therefore further research is needed to achieve classification for continuous mass flows.

2 Introduction

Integrated throughflow mechanical mircofluidic sensors are sensors which are able to attain the mass flow by the mechanical properties of the sensor. Knowing the mass flow is essential to a variety of medical and industrial applications [1]. These sensors also provide information on the physical quantities of the fluid which could potentially be used to determine the fluid inside the tube. However, these are usually not used. The focus of this research is classifying fluids using the data from the mass flow sensor. The sensor used in this research is the integrated mass flow and pressure sensor from IDS. This sensor consist of a Coriolis mass flow sensor and two resistive pressure sensors. The data is collected by generating a data set using the physical relations between the fluids.

The signals from this sensor are used to determine the mass flow but also contain information regarding the physical quantities of the fluid such as the density and viscosity. These parameters can be used to identify the fluid and also estimate contributions of a fluid to a composition [1]. This information can potentially be used in various fields. For example, to determine the quality of crude oil. These oils are essential to the Europe's market and come from several locations such as Russia and Iran. However, when buying these oils, it is important to know the quality. The quality is mainly characterised by the density and sulfur content [10]. Logically, the mass flow sensor would be able to determine the density of the oil and thus give an indication of the quality. Determining the fluidic properties can be solved analytically using the physical relations such as the equation of Hagen-Poisseuille but, these relation only hold under specific circumstances. Machine learning provides a means to bypass these analytical relations and instead learn these relations from data.

Machine learning is a sub-field of artificial intelligence focused on learning and improving mathematical models. They are improved through experience and do no not need any analytical relation instead it approximates the relations from the data. Learning these relations can be performed using three learning methods: Supervised learning uses labeled training data, unsupervised tries to cluster new data and reinforcement learning improves by evaluating the results after execution. This thesis uses supervised learning because of the availability of labeled data and the desired application. Supervised learning provides classification and regression. Models which estimate continuous parameters perform regression while classification is used for discrete outputs. The classification models could be used to identify the fluid inside the sensor while regression can potentially estimate the physical quantities of the fluid.

This research will implement machine learning in a similar fashion to that of human activity recognition (HAR) because this is the closest related field. HAR uses machine learning to classify movement patterns, such as sitting down and walking, using information from various sensors. Performing machine learning with the data from these sensors requires three steps. First, the data needs to be segmented which refers to filtering and windowing the signals to split the signal into several slices with as goal to generate more training data. Secondly, feature extraction is applied to all these slices which extracts the characteristics such as the amplitude or frequencies. Finally, classification models are generated and optimized. Optimization includes selecting features and optimizing the hyperparameters of the model. For this research, the sensor data will be pre-processed to generate data points with every data point having a label referring to the fluid and relevant features which relate to the physical quantities of the fluid. Pre-processing will be done in Matlab while the machine learning algorithms are performed with RapidMiner. The algorithm should learn the relation between the features and the labels and bases the classification on this relation.

Aim of this thesis

Integrating artificial intelligence into the mass flow sensors gives rise to various applications such as estimating the compositions and classifying fluids. This thesis introduces machine learning to the mass flow sensor by using the algorithm to classify the fluids inside the integrated mass flow and pressure sensor from the physical relations between the fluids and the signals.

Note on corona impact

The corona pandemic of 2020 limited this research as the labs at the University of Twente closed down. Therefore, it was impossible acquire data from the sensor, instead generated data is used. The data consists out of six fluids which are generated using the physical effects of the sensor.

3 Theory

This section is split into the theory behind the integrated mass flow and pressure sensor and artificial intelligence. The following section will introduce theoretical concepts on which the current research is based. First, basic properties of the mass flow and pressure sensor will be discussed because the sensor serves as basis for the algorithm. The reader is invited to read the following papers for a more in depth explanation. [3][2][4]. The second part discusses artificial intelligence which explains the principles of machine learning.

3.1 Integrated mass flow and pressure sensor (IMFP)

The IMFP sensor consists of one tube with four different read-out structures. Two of these are used for measuring the displacement and will be referred to as the displacement sensors. the other two are resistive pressure sensors and are referred to as the pressure sensors. An illustration of the device can be seen in figure 1. The figure shows the pressure sensors depicted as the resistive pressure sensor and the displacement sensors at the Coriolis mass flow sensor. The sensor serves three purposes: measuring the mass flow, determining the pressure drop over the tube and determining the density of the fluid. The mass flow is determined by the mechanical displacement of the tube in relation to the Coriolis force [1]. The pressure drop is determined using the two resistive pressure sensors and the density is related to the movement of the tube in one direction. The following section provides an analysis on the relation between the output signals and the physical parameters.

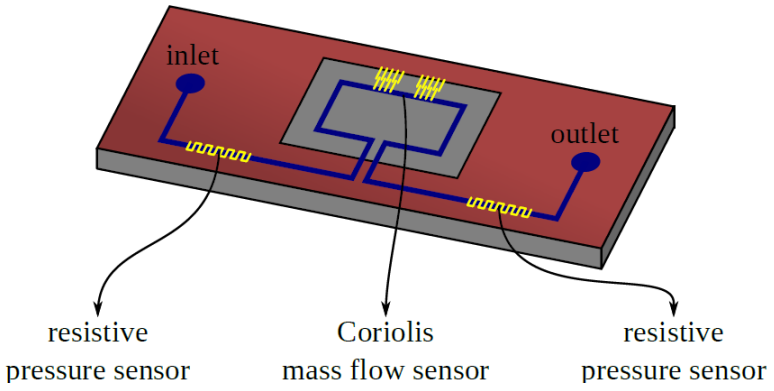


Figure 1: Overview Coriolis mass flow sensor with two pressure sensors[2]

The Coriolis mass flow sensor determines the mass flow by measuring the displacement of the tube. This displacement is influenced by the Coriolis force.

The Coriolis effect is observed when the radius of a rotating object changes. When the mass goes towards the center of rotation, the mass will not follow a straight line to the center. Instead, the Coriolis force will act on the mass perpendicular to straight line. This divergence is referred to as the Coriolis effect and is affecting the tube due to an artificially generated vibration. This vibration is depicted in Figure 2 as the twist mode. Without any mass flow, the mass inside the tube would thus oscillate. However, as the mass has a rotation and moves towards the center of rotation, a Coriolis force will act on the tube. Furthermore, the tube does not rotate but vibrates. As the direction of rotation changes, the direction of the Coriolis force will change thus, resulting in a vibration instead of rotation. This vibration is depicted as the swing mode in Figure 2. The combination of the two signals can be used to measure the strength of the Coriolis force. The twist mode generates a sinusoidal displacement but, the swing mode interferes with another sinusoidal signal. This interference results in a delay between the displacement signals and can thus be observed as a phase shift between the two signals. The combination of the two movements

is shown in Figure 3. By determining the delay, the Coriolis force can be determined which is proportional to the mass flow.

The displacement of the tube is measured using two capacitive combs on the left and right side of the tube. The capacitors can be seen in Figure 2 as S1 and S2. The capacity is inversely proportional to the distance between the combs. Ideally, the capacitance would follow the displacement. However in reality, the is not a perfect sine wave. When the combs cross each other, the capacitance is maximal after which it decreases again thus resulting in a valley.

The capacitors are placed on opposite sides of the twist mode axis. Therefore, the signals will have a phase shift of 180 ° if there is no mass flow. However, this phase shift will change due to the Coriolis force. Therefore, the phase shift of the two displacement signals holds a direct relation to the mass flow Q_m.

$$\Delta\phi \propto Q_m \tag{1}$$

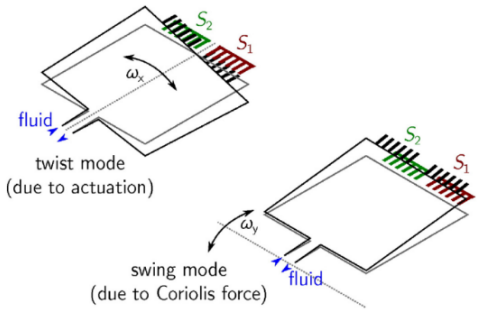


Figure 2: The two vibration modes present in the Coriolis mass flow sensor [2]

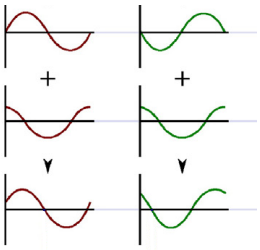


Figure 3: The signals in the Coriolis mass flow sensor. Red: The displacement of the left side of the tube. Green: The displacement of the right side of the tube. Top: Twist mode signal. Middle: swing mode signal. Bottom: Combination of the signals. [4]

The pressure is measured with a resistive pressure sensor. The sensor uses a Wheatstone bridge of which one resistance changes based on the pressure. The resistance is a thin film gold on the top of the tube. The tube is fixed to the silicon, but the fabrication process results in a flat top. This membrane deforms when pressure is exerted. The gold films on the top will therefore stretch and thus change resistance. The Wheatstone bridge transforms this resistance change into a voltage. The difference between inlet and outlet voltage is related to the pressure. [2]

The pressure has a direct relation to the dynamic viscosity of the fluid as follows from the rule of Hagen-Poisseuille.

$$\Delta p = \frac{8\mu L Q_v}{\pi R^4} \tag{2}$$

This equation says that there is a direct relation between pressure drop, dynamic viscosity and volumetric flow rate Q_v. The volumetric flow rate is the rate of volume through the tube and can be expressed as a combination of density and mass flow. This equation holds for straight line flows such as the straight line of the tubes. The volumetric flow rate and pressure difference relate to the dynamic viscosity as follows:

$$\frac{\Delta p}{Q_v} \propto \mu \tag{3}$$

The density is related to the displacement of the tube in the twist mode. The tube is a mass spring system and therefore has a resonance frequency. This resonance frequency is observed in the actuation signal. The resonance frequency in a mass spring system is inversely dependent on the mass. Because the volume of the tube is a constant, the density determines the mass. Therefore, the resonance frequency holds an inverse relation to the density.

$$f \propto \rho^{-1} \tag{4}$$

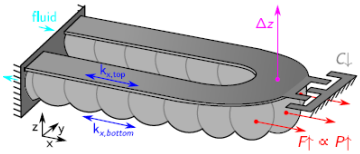


Figure 4: The graphical representation of the tube and the capacitor combs [2]

3.2 Artificial Intelligence

The various relations between fluids and the signals from the IMFP sensor can be used to identify the fluid inside the tube. Artificial intelligence able to to solve complex problems by learning. Therefore, this could be an efficient and reliable approach to classifying the fluids with the sensor's data.

Artificial intelligence can be implemented by either machine learning or deep learning. Machine learning is the traditional approach to artificial intelligence. The performance of the various algorithms is fully understood and can be modified to the needs of the application. However, deep learning outperforms machine learning in most applications. Its main benefit is the automatic feature extraction while machine learning needs the designer to identify the features. The manual extraction of features often limits the ability of the model as human expertise is unable to generate high-level features. Deep learning is capable of designing significant high-level features tailored to each scenario and data type [12]. However, the models are usually more complex than those of machine learning. This makes it difficult to understand the performance and decisions. Therefore, this research will use machine learning.

This research investigates the possible uses of machine learning when used on the data from the IMFP sensor. That is, using the previously filtered information to classify fluids and analysing whether it is possible to estimate the physical quantities of the fluid. There are three algorithms used: Decision tree, k nearest neighbour and naïve Bayes. These three classification algorithms are selected because of their performance in other fields [9]. The next section will explain the principles of classification with hyperparameters. Lastly, a small note on linear regression will be provided to explain the basic principles of parameter estimating.

Supervised and unsupervised learning are two similar approaches to learning. Shortly put, supervised learning uses labeled training data when learning while unsupervised tries to determine probability densities from experience; clustering new data. There are also semi supervised learning algorithms which combine an initial set of supervised learning and continues to learn from new examples. The sensor data provides relations to the fluid thus can be

used as training data. Therefore, supervised learning will be performed.

For supervised learning, the training data needs to be prepared in such a way that the algorithm can distinguish the fluids with the features. The data consists out of features and labels, the labels identify the class and the features the properties. For instance, water has a viscosity of $0.53mPas$ which is a feature of the class water. The algorithm attempts to approximate the mapping function which maps the values of the features to the labels. [9] The data from the IMFP sensor contains four outputs which have various features such as actuation frequency, pressure drop, etc. These features hold a direct relation to the fluidic parameters and can thus be used for classifying fluids. Thus, the training data should consist of the various features related to the physical quantity of the fluid and the fluid itself as label.

Classification and regression are two terms used to describe the output of a model. For classification the output of the model is discrete and can thus only have a finite set of options. For regression, the output is a continuous variable, a number. This thesis is focused on classification of fluid while regression will be used to look for potential research into estimating fluidic parameters and the mass flow. Classification needs labeled data with every label corresponding to one class. The algorithms learns to correlate features with classes. New data will therefore be classified according to its features. E.g. the algorithm could therefore learn that an actuation frequency of 2400 Hz corresponds to water.

Hyperparameters are parameters which change the learning process. Changing the parameters can have a vast impact on the the resulting model and its performance. The hyperparameters are different for every algorithm and must be optimized for maximal performance [6]. Hyperparameters can also be used to prevent overfitting. Overfitting occurs when a model only adheres to the test data and will act wrong for the verification data. The model uses random relations which are not actually there. Hyperparameters can be used to prevent overfitting. [5]

3.3 Machine Learning

Machine learning encompasses various learning algorithms. There are three machine learning algorithms tested: decision tree, k-nearest neighbour and naïve Bayes. Moreover, linear regression will be mentioned as one of the regression methods. The three algorithms will be used to classify the fluids. The implementation will be discussed below.

The algorithms are trained using training data which first needs to be created. To generate training data, two steps need to be performed: segmentation and feature extraction. Segmentation covers the various signal processing techniques such as windowing and filtering the data. Common approaches are frequency filtering with a customized filter and windowing. Windowing refers to splitting the data into several similar data points. By providing more examples, the algorithm can often perform better. The sliding window algorithm is commonly used for segmentation. This method splits the data into equal sizes and can also incorporate overlap between these windows thus creating more data points without influencing the window lenght. The sliding window will be implemented, but without overlap as more than enough data was available.

The feature extraction process refers to selecting and creating features from the windows. These features should provide information regarding the labels, preferably unique for every label. The data generated by the Coriolis mass flow sensor gives various features related to the fluid. The actuation frequency should be extracted as it relates to the density of the fluid. The viscosity can be attained using the pressure drop and mass flow. This relation can be extracted by determining the phase shift between the displacement signals and the difference between the signals from the pressure sensors. Therefore, the phase shift, pressure signal difference and actuation frequency will be provided along with other features.

The machine learning algorithm is provided with data sets from three fluids in which the mass flow and pre-pressure will be varied. The sensor data corresponding to the fluids are generated based on the physical effects of the sensor. The physical quantities of the fluids are shown in Table 1. The density of ethanol, isopropanol and

fluids	density (kg/m^3)	viscosity ($mPas$)
Chloroform	1480	0.53
Ethanol	789	1.1
Hexane	655	0.297
Isopropanol	786	2.4
Methanol	792	0.56
Water	998	1.0

Table 1: The physical quantities of the six fluids

methanol are similar thus making it difficult to distinguish the fluids solely based on the density. However, the viscosity provides a second feature with which the other fluids can be identified.

3.3.1 Decision Tree (DT)

The decision is machine learning algorithm which takes a vector of features such as pressure, mass flow and actuation frequency and returns a class according to the generated decision tree. This machine learning technique is considered to be the best machine learning algorithms regarding performance [9]. The tree consists of nodes and branches in which the node specifies a splitting variable and the branches are the corresponding value ranges. The first node is called the root node and the depth of a tree is specified by the number of nodes from the root node. The last node in every branch is called the leaf node. The leaf nodes specify the output class of the decision tree. For example, the decision tree in Figure 5 shows as root node the density with three leaf nodes stating which fluid belongs to which properties.

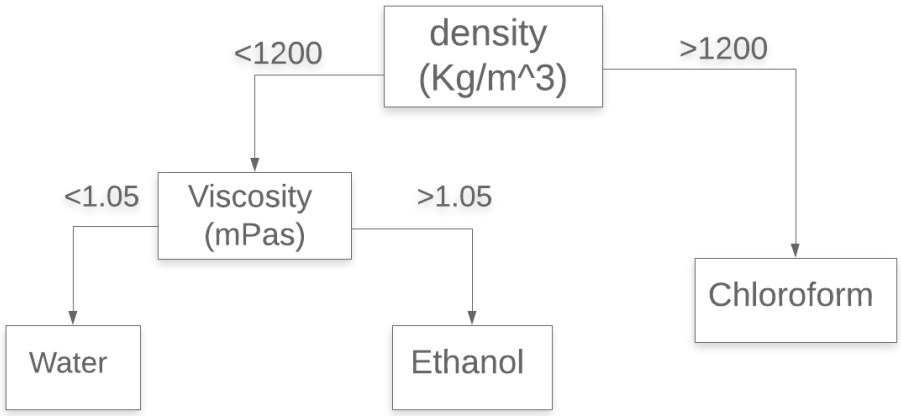


Figure 5: The decision tree identifying fluids based on their physical quantities.

The features generated by IMFP should provide a similar looking decision tree with the density replaced by the actuation frequency and the viscosity replaced by a combination of mass flow and pressure drop. The tree generates the nodes based on the entropy of the attributes. This entropy specifies how much data points are separated by that feature. The feature with the highest entropy is chosen after which the process is repeated. The density is able to distinguish water, hexane and isopropanol clearly which will thus relate to a high entropy for the actuation frequency. Furthermore, the viscosity is capable of distinguishing ethanol, isopropropanol and methanol. However, this relation does require the tree to split both on the mass flow and the pressure drop.

There are several hyperparameters for the decision tree but only the tree depth will be modified. The decision tree is prone to overfitting as it can continue splitting until every single example is classified correctly as long as there are small distinction between the features. Therefore, the model will classify the training data correctly but will not correctly classify new data. The maximal tree depth specifies the maximum number of nodes from the root node and thus limits the capabilities of the decision tree.

The second method to prevent overfitting is pruning. This process removes the clearly irrelevant nodes; nodes which do not hold any relation between label and values [9]. Pruning is usually done after generating the model but can be applied beforehand as well. This method is referred to as pre-pruning and excludes features from becoming splitting variables. This research uses the automatic pruning provided by RapidMiner.

3.3.2 k-nearest neighbor (KNN)

This algorithm classifies new data based on the classes of the data points which have the most similar features. This algorithm is a passive algorithm which means it does not have a learning procedure. Instead, KNN classifies the data during execution. The algorithm constructs a data space with the features used as axis. The data points are located according to their respective feature values. This can be seen in Figure 6. The plot shows three classes which seem to be separable but do show overlap. KNN uses this data space to perform classification on new data by looking at the features of the data. It places the new data point in the data space and determines which data points are closest to it. The k specifies how many of its closest neighbours are considered. The decision is based on which class occurs most often in its neighbours.

To illustrate this process, a gray dot is included in Figure 6. The algorithm uses a k of three to classify the new data. The three closest neighbours are represented by the black line. The classes of these data points are red therefore, the algorithm will classify the gray dot as red.

The algorithm works best for data with a small spread. The classes in Figure 6 show a relatively large distribution causing overlap between the classes. This results in faulty classification. For example, if the left most red dot would be new data instead of example data, the algorithm would classify it as green. Therefore, the accuracy and resolution of the features play a significant role in the performance of the algorithm.

Furthermore, there are two hyper parameters important for k nearest neighbour. The first is k which specifies the number of names taken into consideration. Furthermore, the distance between the neighbours and the new data point can be taken into account, this is called weighted voting. These two hyperparameters change the behaviour of the model.

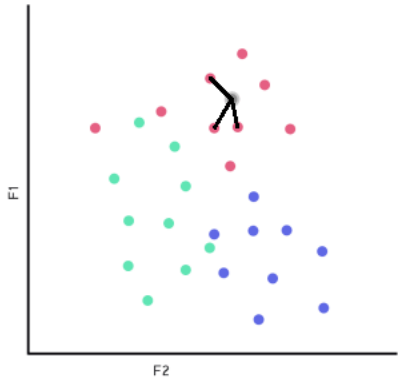


Figure 6: Data space with three classes and two features (f1 & f2). The gray dot represents new data and the black lines its closest neighbours.[8]

3.3.3 Naïve Bayes (NB)

Naïve Bayes is a probabilistic classification algorithm based on Bayes rule. The model learns dependencies between features and classes and outputs probabilities of new data belonging to the different classes. The algorithm bases the probabilities on the recurrence of a certain feature in a class. I.e. if class A only contains an actuation frequency of $2500Hz$, then the algorithm will assign new data with an actuation frequency of $2500Hz$ a high probability of belonging to class A. However, the decision is based on the complete probability based on all features.

The mathematical foundation is probability theory and Bayes' rule which states:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{5}$$

This is translated to features as follows. The first part of the equation states the probability that unseen data B belongs to class A. Bayes' rule states that this probability can be rewritten as the probability that the features of B are present in class A times the probability of class A occurring divided by the probability that the features of B occur. [9]

The probabilities on the right side of the equation can be calculated. The probability of class A having features B is determined by counting how often the features occur in class A. For instance, the class water might have an actuation frequency of $2500Hz$. If there is noise, the entire water class will have this feature thus resulting in a probability of 1 that the feature occurs in the class. The probability that class A occurs is determined by dividing the number of occurrences of A divided by the total number of data points. Lastly, the probability that features B occur is determined by counting how often the value occurs divided by the total number of data points. E.g. a data set of 100 data points of which 20 water will give a probability of $\frac{1}{5}$ for $P(A)$ and if only water has an actuation frequency of $2500Hz$, the probability of the feature occurring is $\frac{20}{100} = \frac{1}{5}$. Therefore, a new data point with this freuquency will have a probability of $\frac{1 \times .2}{.2} = 1$. Therefore, naïve Bayes works best when the features are unique for every class.

Naïve Bayes can work classification of fluids when the features are unique for every class. Furthermore, the classification is based on the number of occurrences in the data. Therefore, the performance will depend on whether the data set has an equal number of data points for every class.

3.3.4 Linear regression

Linear regression models the relation between two dependent variables. The relation is determined by a linear prediction function which generates a linear function of the form $y = aX + b$. The example data provides relations between features X and output y which are used to tweak the parameters a and b. This can be used to estimate linear relations such as the relation between the actuation frequency and the density.

When the model contains multiple features, multivariate linear regression needs to be applied. This model can predict relation between multiple features and the output. The output y becomes a function of multiple features $\overrightarrow{X}$. The complete function becomes $y = \overrightarrow{a}\overrightarrow{X} + \overrightarrow{b}$. The size of the vectors is equal to the number of features. This becomes applicable when estimating the viscosity which is based on the combination of mass flow and pressure.

4 Method

This section will describe the various steps needed to perform machine learning on the data from the IMFP sensor. This includes data analysis, data preparation (segmentation), feature extraction and the learning process. The data preparation is performed in Matlab R2020A and the machine learning with RapidMiner. Matlab is a commonly used platform for digital signal processing and other mathematical analysis. RapidMiner is one of the various frameworks on which you can perform machine learning.

This research uses data from the four read-out structures on the IMFP sensor. To recap on this sensor, the sensor provides four signals. Two for the inlet and outlet pressure and two for the displacement of the tube. The output of the sensors is connected to a DAC with a sampling frequency of 250.000 Hz. This makes it possible to distinguish between signals up to 125KHz [11]. The following parameters are changed between simulations.

1. Mass flow
2. Pressure
3. Fluid

There are six different data sets generated for six different fluids: Chloroform, ethanol, hexane, isopropanol, methanol and water. The mass flow was varied from zero g/h till eight g/h with steps of one gram per hour. Lastly, the pre-pressure is varied between three bar, six bar and nine bar. For every combination, one second is simulated providing 250.000 samples. From this data, three features must be extracted: Pressure drop, actuation frequency and phase shift.

The relation between these features and the fluid is discussed in the theory section 3. Moreover, the higher harmonic contains information on the fluid. As previously described, the signal will have a valley whenever the capacitor combs cross the chip which can be observed in the frequency spectrum by a second harmonic. This phenomenon is visualised by the valley between the two peaks on both signals shown in Figure 7.

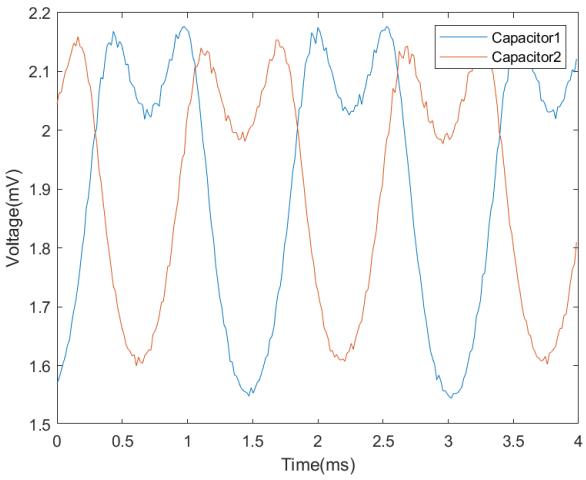


Figure 7: The measured voltages related to the signals generated by the two displacement sensors for ethanol with a three bar pressure and a mass flow of four grams per hour

4.1 Pre-Processing

To apply machine learning on the data, segmentation and feature extraction will be performed. First segmentation will divide the data into several windows after which features are extracted from the windows. The following features will be extracted from the data: actuation frequency, voltage difference between inlet and outlet pressure sensors, phase shift between the capacitor signals and the magnitudes of the first and second harmonic.

4.1.1 Segmentation

Segmentation focuses on both the filtering windowing of the signals. The original signal is one second long and sampled at a sampling frequency of $250.000Hz$. The signals are stored in excel files and read into the workspace of Matlab using the readtable function and transformed into an array using table2array. The result is a data set represented as a matrix of 27 by 250.000 by 4 for every fluid. 27 due to the three pre-pressures and the 8 different mass flows and 4 because of the four different sensors. This process is executed by the readfiles.m function.

These signals are windowed into even length windows using the reshape function from Matlab [**reshape**]. Every window will generate one set of features. The reshape function reshapes a matrix into the dimensions given as input. The function was made generic by using a variable called "windows" which controls the number of generated windows. Originally, 100 windows were prepared resulting in a 27 by 2500 by 100 by 4 matrix for every fluid. The size of 100 windows was chosen based on a balance between samples per window and available training data. Figure 8 shows an example of this window for the capacitors.

There is no general filtering applied to the data. Instead, the filtering is executed during the feature extraction. This increases the accuracy and resolution of the features as the filter can be modified to increase the accuracy of the feature.

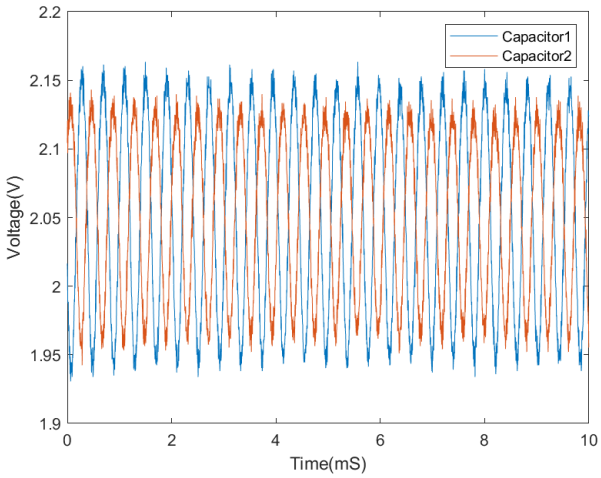


Figure 8: One example window of the capacitances for water with nine bar pre-pressure and 4 grams per hour mass flow.

4.1.2 Pressure analysis

The first feature to be extracted is the pressure drop. This is attained by determining the voltage difference between the two pressure sensors. The signals from the pressure sensors are shown in Figure 9. The signal contains large amounts of noise. To filter this noise, the DC signal is taken. For efficiency, the mean function is used which can take the average along one dimension of a matrix. The result is a 27 by 100 matrix containing the voltage difference for every mass flow, pre-pressure and window.

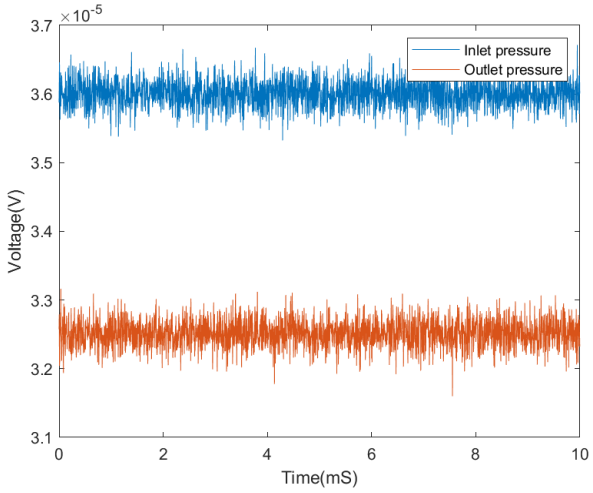


Figure 9: The two pressure signals extracted from one window. The window is taken from the water data set with nine bar pre-pressure and a mass flow of 4 g/h

4.1.3 Frequency analysis

This section describes how to extract the features from the frequency domain of the signals. The signals will be transformed using the Fourier transform. Only the two displacement sensors provide information in the frequency domain. The following four features are extracted from the two displacement signals:

- Actuation frequency
- Magnitude actuation frequency
- Second harmonic frequency
- Second harmonic magnitude

The signals are transformed using the fast Fourier transform function from Matlab, but first the DC value is removed as it causes a spike in the frequency domain. Removing the DC value is done using the mean function and subtracting the mean from the data. The fast Fourier transform is applied to the remaining signal. This results in a frequency signal ranging from 0 Hz till 250kHz. However, the signal is mirrored around 125 KHz [11]. Therefore, only the first half of the signal is analysed. The absolute value of the frequency signal is shown in Figure 10. The figure shows a strong peak at $2600Hz$. The actuation frequency for the other fluids is observed between $2400Hz$ and $2600Hz$. The second harmonic is on twice the actuation frequency therefore, the signal could be filtered from $5200Hz$. However, faulty data can easily be removed in a later stage therefore, a frequency range of $0 - 20KHz$ is used.

To automatically attain the actuation frequency, either the max function or the findpeaks function can be used. the max function returns the maximum from the entire signal while findpeaks will return all the maximums in the signal. The benefit of the findpeaks function is that it can be used to find the frequencies and magnitudes of higher harmonics. The output of the findpeaks algorithm can be altered using various parameters. This research uses three parameters specified as MinPeakDistance=10, SortStr=descend and Npeaks=2. MinPeakDistance specifies the minimum number of samples between peaks which is used to prevent similar frequencies being considered twice. SortStr determines whether the peaks are returned from maximum magnitude to lowest magnitude or lowest magnitude to highest magnitude. Lastly, Npeaks specifies the number of peaks considered. There are two frequencies which should be returned: the actuation frequency and its second harmonic.

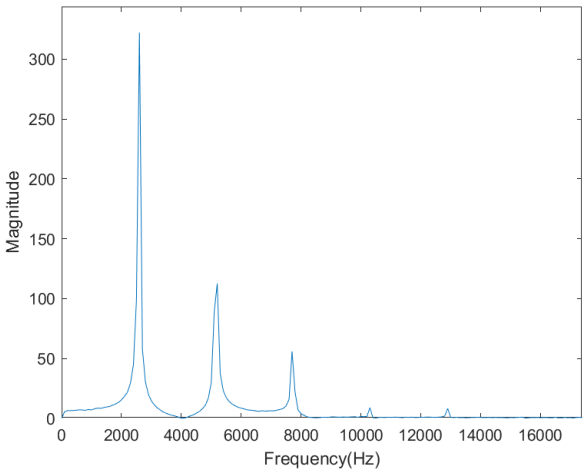


Figure 10: The frequency signal for ethanol at three bar pre-pressure and a mass flow of four g/h.

The function returns the indexes of both the magnitudes and the frequencies from the two displacement signals.[7] To obtain the frequency, an array is constructed with the frequencies corresponding to the indexes. The frequency of the second harmonic is also returned. This does not necessarily provide information but can be used to validate the features.

Unfortunately, this method only has a resolution of 100 Hz. The Fft of 2500 samples gives an output of 2500 samples which are equally divided over the entire frequency spectrum. The frequency spectrum ranges from 0 till 250.000 Hz (the sampling frequency). Therefore, every sample corresponds to $250.000/2500 = 100$ Hz. To increase the resolution, either the window size needs to be increased or the data needs to be resampled at a higher frequency. The first method is applied by decreasing the number of windows to 25.

The execution is performed by a Matlab function which takes the windowed data and returns both magnitudes and frequencies in two matrices of 27 by 100 by 2 by 2. Thus, the first and second harmonic frequencies and the respective magnitudes from the two displacement signals for every window, mass flow and pre-pressure.

4.1.4 Phase shift

Lastly, the phase shift between the two displacement signals will be extracted. There are several methods to extract the phase shift. The Fourier transform provides an angle at every frequency. The angle from one displacement signal can be extracted by taking the angle at the actuation frequency. By doing this for both signals and subtracting them, the phase shift can be determined. The other method is cross correlation, a commonly used digital signal analysis technique [11]. This research uses cross correlation as the Fourier transform was too inaccurate.

The phase shift is defined by $\Delta\phi$ which is the phase shift necessary for two sinusoids to be the same. This can be mathematically written as.

$$sin(\omega t + \phi_1) = sin(\omega t + \phi_2 + \Delta\phi) \tag{6}$$

This equation has two signals with equal phases due to the extra phase shift introduced by $\Delta\phi$. The phase shift is given as $\Delta\phi = \phi_1 - \phi_2$.

Cross correlation shifts two signals over each other and determines the correlation between them. The displacement signals are discrete therefore, the signals are shifted across each other in discrete steps. If the signals are identical, the output is maximal while a 180° shifted signal will result in a minimum. The cross correlation of the two displacement signals will generate a sine like pattern due to the periodic behaviour of the signals. However,

the magnitudes decrease when going further from the centre as can be seen in Figure 11. This decrease is due to the the length of the signal which is 2500 samples. The signals are shifted across each other but are not extended. Therefore, the shifted signal overlaps with zero and is thus not correlated.

The number of samples shifted corresponding to the maximum correlation is a direct indication of the phase shift. The index of the maximum corresponds to the amount of samples the signal is shifted. From this sample lag, the phase shift can be determined from the number of samples per period. The number of samples per period is calculated using the actuation frequency. The time per period is one over the actuation frequency $Ts = 1/f_{\text{act}}$. The sampling frequency is the number of samples per second. Together, the samples per period can be determined. The example equation below gives the samples per period for a fluid with actuation frequency $2600Hz$.

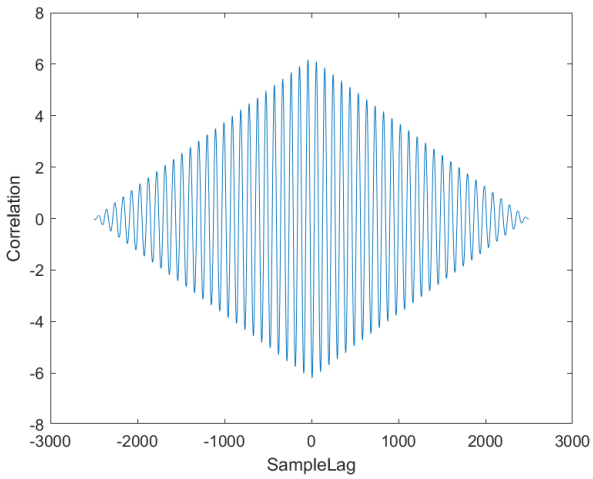


Figure 11: The cross correlation of one ethanol window

$$Sn = Ts \times Fs \tag{7}$$

$$Sn = \frac{1}{2600} \times 250000 Sn \approx 96.1 \tag{8}$$

Sn gives the samples per period. Dividing the sample lag by the samples per period gives the phase shift.

$$\Delta\phi = \frac{Samplelag}{Sn} \tag{9}$$

However, the resolution is limited as the signals can only be shifted by one sample. Therefore, the resolution for the example fluid is:

$$Resolution = 1/Sn \tag{10}$$

$$Resolution = 1/96.1 \times 2\pi \approx 0.065 Rad \tag{11}$$

To increase the accuracy, two methods are applied. First, the other peaks shown in Figure 11 are used and averaged. The second modification is upsampling. The limiting factor is the number of samples per period. Upsampling will resample the signal at a higher sampling frequency thus generating more samples per signal. This research used an upsampling factor of four thus generating a signal of 10.000 samples out of the original 2500 samples [11]. However, this method introduces distortion at the beginning and the end of the signal. This does not influence the detected phase shift as there is no correlation between the distorted part of the signal and the other signal. With

four times as many samples, four times as many steps can be taken and thus the resolution is increased by a factor 4.

Again a function is created which takes the windowed fluid data and returns a 27 by 100 matrix providing the phase shifts between the capacitor signals for every pre-pressure, mass flow and window.

The complete set of features is written into a csv file using the writetable function from Matlab. The table inside the csv file contains thirteen columns and 2700 rows for every fluid. The columns are the labels and features of the data. They are ordered as follows: Fluid, Mass flow, Pressure, Actuation frequency displacement sensor 1, Actuation frequency displacement sensor 2, Voltage difference from the pressure sensors, Phase shift, Magnitude actuation frequency displacement sensor 1, Magnitude actuation frequency displacement sensor 2, Second harmonic frequency displacement sensor 1, Second harmonic frequency displacement sensor 1, Magnitude second harmonic frequency displacement sensor 1 and Magnitude second harmonic frequency displacement sensor 1.

4.2 Machine Learning

The generated data set can now be used for machine learning. The algorithms discussed in section 3 will be applied and tested. Furthermore, the performance will be maximized by optimizing the hyperparameters, feature selection and normalization. First the capabilities of RapidMiner will be discussed and the performance measure after which the implementation of the classification algorithms will be provided.

This research uses RapidMiner as machine learning framework. RapidMiner is a free to use data science platform providing "processes" which can be used for machine learning and big-data analysis. There are three programs available: RapidMiner Go (free), RapidMiner Studio (business) and the Educational Program. This research uses the Educational Program which provides access to all the functions of RapidMiner Studio.

The platform provides hundreds of pre-made processes for machine learning and a graphical user interface. These processes are built-in machine learning algorithms such as decision tree and data preparation processes such as feature selection and data filtering. The results of the processes are graphically and textually presented. The downside to RapidMiner is that it is not as versatile as other platforms and is not suitable for creating a commercial product as only the pre-defined algorithms can be used instead of specifically designed algorithms. Nevertheless, this research uses RapidMiner as it provides an intuitive way to test various algorithms.

4.2.1 Performance Measure

The performance measure determines the performance of the model based on the classification error. To measure the performance, unseen data needs to be provided and a performance measure must be set. This unseen data is often a subset of the data set and is referred to as the validation data. The performance measure specifies the influence of wrongly classifying a data point based on the desired goal of the application. Creating the validation set will first be explained, after which the the three performance measures will be explained.

The validation data is often a subset of the complete data set created by either hold-out or cross validation. The hold-out method simply takes a part of the complete data and separates it for validation. Usually 20 percent of the total data is used to validate the performance. Obviously, this process results in worse performance because not all the information can be used. Moreover, only a subset is used to measure the performance. Cross validation prevents both of these problems. It partitions the data into k different subsets of equal size. The program will then train the model using k-1 subsets and validate the model using the left data set. This process is repeated for every subset thus, using all the available data for training and validating. With every iteration, the number of correct and incorrect predictions is stored. The total performance is given by the average performance over each iteration.

There are three performance measures used in this research, namely the accuracy, precision and recall. Accuracy gives a performance based on the percentage of correctly classified data; $accuracy = \frac{TP+TN}{TP+FP+TN+FN}$. For example,

four fluids to classify and the validation set consists out of 100 data points equally divided over the four fluids. The model classifies 80 out of 100 data points correctly then the accuracy is 80% ($\frac{80}{100}$). This states that the model gives a correct prediction 80% of the time.

Class recall is specified for every class by the number of correctly classified divided by the total number of data points in the class. Written in the from of true positive and false negative as $Recall = \frac{TP}{FN+TP}$. E.g. a validation set has 100 water points of which 80 are classified as water and 20 as ethanol. The recall is then $80\% = \frac{80}{80+20}$

Lastly the precision which specifies the accuracy for every individual fluid. The percentage reflects how sure the prediction is with respect to every fluid. The accuracy is an average of the precision, but the precision can vary between fluids. The precision is given as $precision = \frac{TP}{TP+FP}$

These methods only work for data sets in which every class has an equal weighing. However, when it becomes more important to classify one class correctly over the others, different performance measures need to be used. However, this will not be used for this research as every fluid will have the same importance. The cross correlation process in RapidMiner provides an automatic method to determine both the class recall, the precision and the accuracy.

4.2.2 Rapid miner

There are three machine learning algorithms used, namely the decision tree, naïve Bayes and k-nearest neighbour. This section describes how to perform the three different classification algorithms with the data. Furthermore, the tuning of these model will be explained.

The original data set contains 16200 data points with three labels and ten features. There are 2700 data points generated per fluid with eight mass flows, three pre-pressures and 100 windows. The following ten attributes are provided.

1. Actuation frequency displacement sensor 1
2. Actuation frequency displacement sensor 2
3. Voltage difference between pressure sensors
4. Phase shift
5. Magnitude actuation frequency displacement sensor 1
6. Magnitude actuation frequency displacement sensor 2
7. Second harmonic frequency displacement sensor 1
8. Second harmonic frequency displacement sensor 2
9. Magnitude second harmonic displacement sensor 1
10. Magnitude second harmonic displacement sensor 2

The performance of three classification algorithms is completely dependent on the data set. Therefore, various changes will be made to the data set to increase the performance. The first step is to ensure no faulty data is present. This will be achieved with the filtering process which excludes data points with features out of the correct range. Initially the data contains the mass flow and pressure features but these are actually labels. Therefore, these features will be removed from the data set. Next, normalization will be applied to subsets of features which transforms the values into new ranges. Lastly, feature selection will be applied.

The data was filtered with the second harmonic frequency. Second harmonic frequencies outside of the range of $4500Hz$ and $5300Hz$ are filtered as will be explained in the result section. This results in a data set of 15614 total data points thus, 586 filtered data points.

After filtering faulty data points, the mass flow and pre-pressure features need to removed. Removing features can be performed using the attribute selection which allows the user to use a selection of features

Normalization will be applied to a selection of the features. The most important feature to normalize is the voltage difference between the pressure sensors. These features are in the range of 10^-5 and are therefore difficult to distinguish. Normalization transforms them back to a specific range. The normalization procedure used in this research is the z-transform. The z-transform transforms the data into a new data set with the mean equal to zero and a variance of one. It is applied to all the features but only the voltage difference seems to impact the performance.

Feature selection is applied to determine the influence of the features on the performance of the algorithm. The features are selected using the attribute selection process. The selection is based on the physical relation between the feature and the fluid thus selecting the phase shift, voltage drop and actuation frequency as these are related to the viscosity and density of the fluid.

Hyperparameter optimization will be applied with every step to increase the performance of the decision tree and k nearest neighbour. This method uses the Optimize Parameter process from RapidMiner. This process provides a method to iterate over the values of the hyperparameter and simultaneously determine the performance of the model. For the decision tree, the maximal tree depth is varied from 1 till 100. For k-nearest neighbour, two parameters are tuned. The first parameter is the k which states how many neighbours should be considered and the second parameter is whether the distance to the neighbours should be weighted. The k will be varied from 1 to 100 with steps of 5 and the weighted votes will be altered with every step between on and off. Naïve Bayes does not have any hyperparameters to tune. hyperparameter optimization is tuned by determining the performance of the models which is based on the output of the cross validation process.

The changes influence the performance of the model. The performance of the models is verified with each change using cross validation. The changes are shown in Figure 12. The results are presented in the next section.

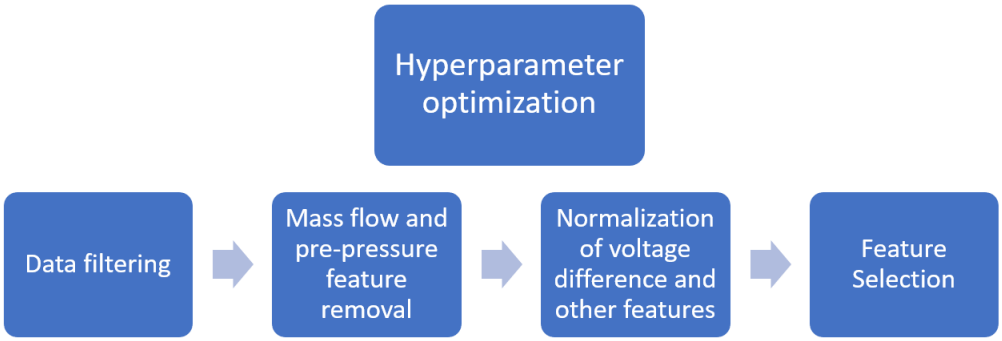


Figure 12: Machine learning performance measurement process.

5 Results

This section presents the results of both the pre-processing and machine learning. The results are produced according to the methods described in the method section. The methods are applied to a generated data set. The data set is generated from a the physical effects of the sensor. The pre-processing section presents the results of both the segmentation and the feature extraction processes. The machine learning section will describe the performance of the various algorithms and the influence of the optimization techniques.

5.1 Pre-processing

The pre-processing consists of the segmentation of the data and feature extraction. The segmentation splits the data into windows from which the features are extracted. There are six fluids used with the mass flow and pre-pressure varied. Every fluid, mass flow and pre-pressure combination has one second worth of data which is split into 100 equally sized windows of 2500 samples. Ten features are extracted from these windows with three functions. The functions give the voltage difference between the pressure sensors, the frequency information and the phase shift. This section presents the results from these three functions.

5.1.1 Pressure results

The first feature is the voltage difference between the inlet and outlet pressure sensors. Figure 13 shows the average voltage drop between the pressure sensors. The average is calculated by taking the average voltage difference over all the windows. Changing the pre-pressure only influences the observed voltage difference by a absolute change of $10^{-8}V$. The standard deviation of voltage difference regarding pre-pressure is approximately $5 \times 10^{-9}V$ for every fluid. Thus, the pre-pressure has almost no influence on the voltage difference. The standard deviation between windows is on average $6 \times 10^{-9}V$ for all the fluids and mass flows. The observed voltage difference is thus only marginally influenced by the pre-pressure and does not vary significantly between windows.

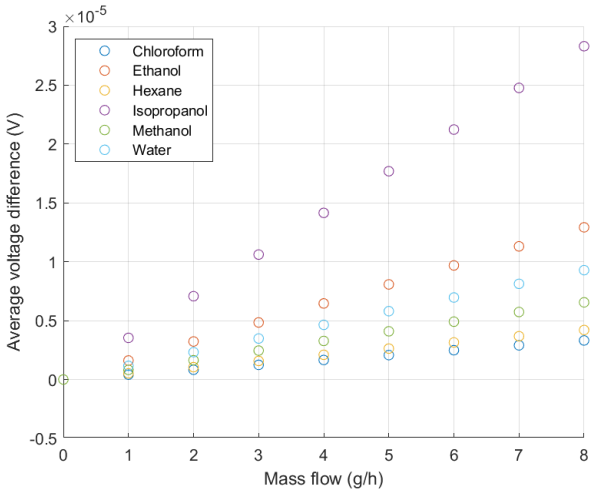


Figure 13: The average voltage difference between inlet and outlet pressure sensor for the eight mass flows and six fluids.

5.1.2 Frequency results

The frequency analysis provides eight features which are extracted from the two displacement sensors. The following four features are extracted from the data for both sensors:

- Actuation frequency
- Magnitude actuation frequency
- Second harmonic frequency
- Second harmonic magnitude

The calculated actuation frequency ranges from $2300Hz$ till $2600Hz$. The distribution per fluid can be seen in Figure 14. The actuation frequency of ethanol, hexane, isopropanol and methanol are always $2600Hz$ while chloroform has an actuation frequency of $2300Hz$ and water either $2400Hz$ or $2500Hz$. Both displacement sensors give these distributions. However, the frequencies show abnormalities with several data points having actuation frequencies at least $500Hz$ from the expected range of $2300 - 2600Hz$.

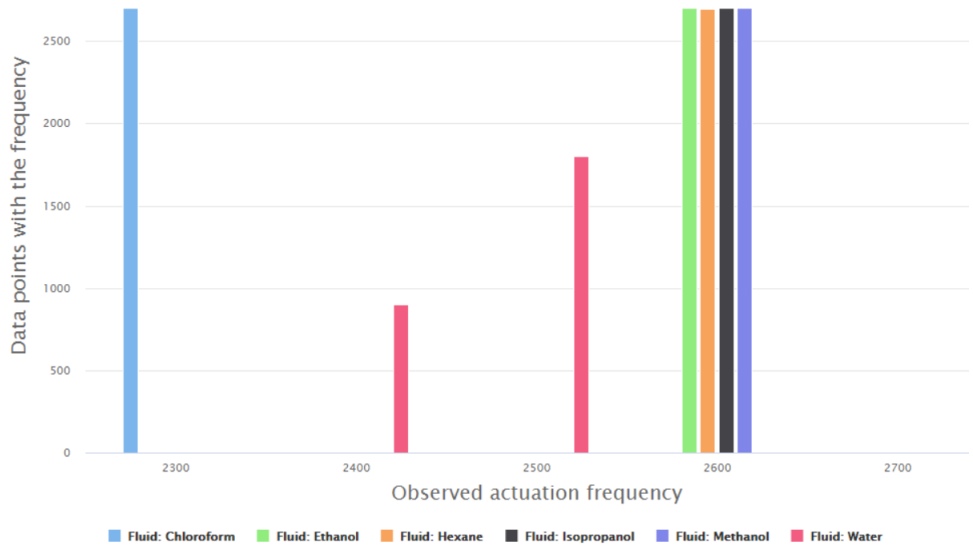
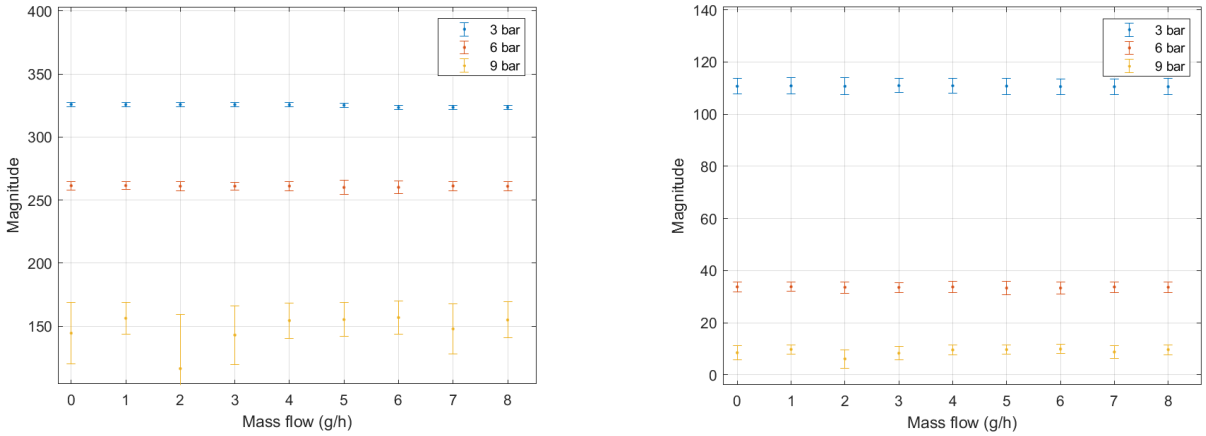


Figure 14: The distribution of actuation frequencies from displacement sensor one for every fluid.

The second harmonic frequency ranges from $4500Hz$ to $5300Hz$. The distribution can be observed in appendix Figure 17. This figure shows the number of data points with the specified second harmonic frequency for every fluid. Ethanol, hexane and isopropanol have a frequency of $5200Hz$, water at $4900Hz$ and chloroform at $4500/4600Hz$ The second harmonic frequency does show more unexpected results with 122 data points outside of the range of $4500 - 5300Hz$ for the first displacement sensor and 560 data points for the second displacement sensor. These abnormalities are at least $500Hz$ outside the range and are thus very likely false data points.

The magnitude of the actuation frequency for displacement sensor 1 is plotted in Figure 15a. The figure shows the average magnitude in an arbitrary unit corresponding to the actuation frequency together with the standard deviation. The unit is arbitrary as machine learning will classify according to the relative values. From the figure it becomes apparent that the magnitude of the actuation frequency decreases with more pre-pressure and the standard deviation increases. However, the mass flow does not seem to influence the calculated magnitude.

The second dependency of the magnitude is the fluid. For 3 bar, methanol, ethanol, and isopropanol all show the same magnitudes of 325, chloroform and hexane are slightly lower with an average magnitude of 296 while



(a) Magnitudes of the actuation frequency

(b) Second harmonic magnitude

Figure 15: The average observed magnitude of ethanol with the error bars representing the standard deviation.

water has a significantly lower magnitude of 233. The relative difference between magnitude stays constant for the different pre-pressures.

The magnitude of the second harmonic can be seen in Figure 15b. The figure shows the average magnitude corresponding to the actuation frequency together with the standard deviation. The absolute standard deviation stays the same for different pre-pressures. However, the standard deviation relative to the average magnitude increases when increasing the pre-pressure.

The relation between pre-pressure and magnitude is the same for the actuation frequency and the second harmonic frequency. Moreover, similar to the actuation frequency, the magnitude is divided into the same fluid sets with chloroform, hexane and water having a lower magnitude.

5.1.3 Phase shift

The phase shift is created by up sampling and cross correlating the displacement signals. Figure 16 shows the calculated phase shift for chloroform with the standard deviation over the windows.

The calculated phase shift shows a relation between mass flow with a slight upwards trend when increasing the mass flow. However, the phase shifts are not unique for every mass flow. This is apparent from the overlap between phase shift and mass flows. Moreover, the calculated phase shift deviates largely between windows as is shown by the standard deviation. This standard deviation decreases with an increase in pre-pressure. Moreover, increasing the pre-pressure decreases the calculated phase shift.

5.2 Machine learning

The generated features are used for the three classification algorithms. This section presents the performance of these algorithms including the change in performance based on the feature modification.

The overall accuracy of every implementation is given in the table below. First the accuracy with all the features is given. Then the mass flow and pre-pressure features are removed. Normalization is applied to the voltage difference for the decision tree and naïve Bayes but for all features when using k nearest neigbhour.

5.2.1 Decision tree

The original decision tree was generated with all the features including the pre-pressure and mass flow features with which it achieved an accuracy of 76.08%. The cross validation results are given in Table 3. The table shows the

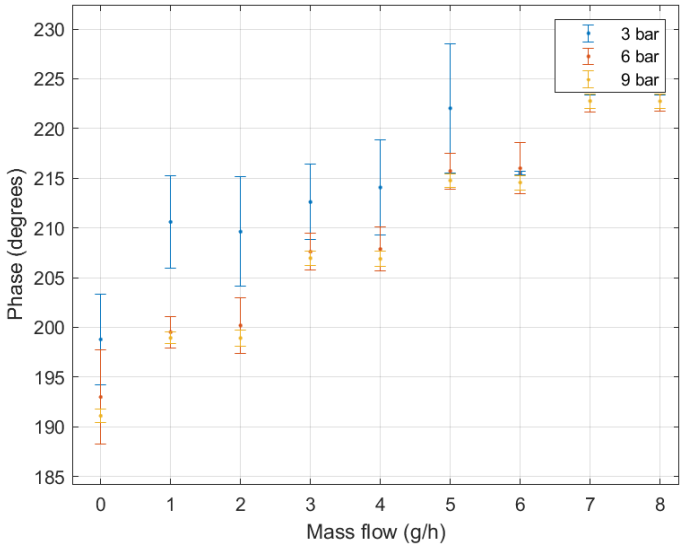


Figure 16: The average phase shift calculated by cross correlation with the error bars representing the standard deviation.

	Original	MF and pre-pressure removed	Normalization	Feature selection
Decision tree	76.08	76.08	95	92.79
Naïve Bayes	65.98	66.18	78.11	65.17
k-nearest neighbours	62.47	62.47	92.96	88.32

Table 2: The accuracy in percentage of the machine learning algorithms with the different feature modifications.

predictions with the given test data. This DT uses the actuation frequency, second harmonic frequency, voltage difference and phase shift to classify the fluids. It is able to distinguish chloroform, hexane and water using the frequencies. Hereafter, the algorithm tries to split the other three using the phase shift and voltage difference but cannot classify all data points correctly.

	true Chloroform	true Ethanol	true Hexane	true isopropanol	true Methanol	true Water	class precision
Pred. Chloroform	2584	0	0	0	0	0	100
Pred. Ethanol	0	2342	0	1022	2343	0	41.04
Pred. Hexane	0	0	2583	0	0	0	100
Pred. Isopropanol	0	0	0	1472	0	0	100
Pred. Methanol	0	261	0	109	260	0	41.27
Pred. Water	0	0	0	0	0	2641	100
class recall	100.0	89.97	100.0	56.55	9.99	100	

Table 3: The performance of a decision tree given by cross validation. The rows represent the predictions and the columns the actual class.

The decision tree generated after removing the mass flow and pre-pressure gives a similar accuracy of 76.08% which is achieved by replacing the mass flow node by the phase shift. The previous DT used the mass flow to make a distinction between ethanol and isopropanol. This DT is able to make the same distinction using the phase shift. The phase shift is not an accurate representation of the mass flow which is the causes for the wrongly classifieid isopropanol data points.

Normalizing the voltage significantly increases the performance up to an accuracy of 95%. The optimal tree

depth is reached at a depth of 15 after which the accuracy remains constant. The decision tree is able to classify 100% of the water and chloroform purely based on the actuation frequency which is unique for both fluids. Next, hexane is split from the rest of the data using the second harmonic frequency. Again this feature is unique for hexane. Lastly, isopropanol, methanol and ethanol are classified using the remaining features. The resulting decision tree is complicated as can be seen in appendix Figure 18. The tree is able to classify chloroform, hexane, and water perfectly while ethanol has a precision of 89.85%, isopropanol a precision of 90.22% and methanol a precision of 94.04%.

Applying the feature selection causes only a slight decrease in performance resulting in an accuracy of 93.79%. Therefore, the decision tree is able to use the priorly defined relation to classify the fluids.

5.2.2 Naïve Bayes

The original model generated with only filtering the data showed an accuracy of 65.98%. This slightly lower accuracy is because not all features contribute to a better accuracy. Naïve Bayes uses probability distributions for classification. Without modifying the features, there are twelve features which contribute to the decision. The mass flow, pre-pressure and phase shift do not relate to a specific feature. They hold an equal distribution for every fluid and thus do not help the algorithm.

The first modification is the removal of the mass flow and pre-pressure attributes which provides a similar result of 66.18%. The precision and recall is given in Table 4. The table shows that the model confuses ethanol, isopropanol and methanol.

	Precision(%)	Recall(%)
Chloroform	100	100
Ethanol	31.93	28.16
Hexane	100	100
Isopropanol	32.35	37.11
Methanol	32.79	31.85
Water	100	99.96

Table 4: The precision and recall of naïve Bayes.

The performance increases to 78.11% when applying normalization on the voltage difference. This increases the accuracy and recall of ethanol, isopropanol and methanol.

The feature selection is applied leaving only the actuation frequency, voltage difference between pressure sensors and phase shift which gives an accuracy of 65.17%. The main decrease is the inability to predict hexane which is always classified as methanol thus leading to a recall and precision of 0%.

5.2.3 k-nearest neighbours

K-nn achieves an accuracy of 62.47% without modifying any features and when removing the mass flow and pre-pressure from the list of features. Moreover, the hyperparameters show similar influence for both implementations. The accuracy of the model shows an upwards trend when increasing the number of neighbours. The accuracy goes from 51.4% for k equal to 1 and goes up to 62.47% when increasing the number of neighbours. Not weighting the distance positively impacts the accuracy of the model as is shown in appendix Figure 19. The confusion matrix is similar to naïve Bayes as it is unable to distinguish ethanol, isopropanol and methanol from each other. However, the class recall of water is 30% as it is also predicted to be methanol.

Normalization of the voltage drop increases the accuracy to 74.48% while normalizing all the features results in an accuracy of 92.96%. The main improvement is the precision for ethanol, isopropanol and methanol which is at least 80% for all fluids. In both implementations, the maximum accuracy is reached after k is equal to 5 and not weighting the votes.

Selecting the actuation frequency, phase shift and voltage drop results in an accuracy of 88.32% for a k of 1 and not weighting the distance. The accuracy decreases when increasing the number of neighbours while the weighing the votes does not influence the accuracy. The Confusion matrix is given in Table 5. It shows that chloroform and water can be separated by the algorithm while, methanol is the most difficult to classify with the lowest recall and precision while the other three fluids have a precision and accuracy around 85%.

	true Chloroform	true Ethanol	true Hexane	true isopropanol	true Methanol	true Water	class precision (%)
Pred. Chloroform	2584	0	0	0	0	0	100
Pred. Ethanol	0	2125	88	171	224	0	81.48
Pred. Hexane	0	74	2185	31	254	0	86.19
Pred. Isopropanol	0	150	47	2254	130	0	87.33
Pred. Methanol	0	254	263	147	2004	0	75.11
Pred. Water	0	0	0	0	0	2641	100
class recall (%)	100	81.64	84.59	86.59	76.99	100	

Table 5: The results of cross validation for k-nn with only the voltage difference between pressure sensors and actuation frequency. The rows represent the predicted label and the columns the actual label.

6 Discussion

This section will discuss the results presented in the previous section. This includes the results from the pre-processing functions and the performance of the various machine learning algorithms. Furthermore, improvements will be presented together with an outlook on future research potential.

6.1 Interpretation of the results

First the results will be discussed along with their implication of the classification of fluids. Hereafter, the performance of the machine learning algorithms will be discussed including the reasons for the varying performance.

6.1.1 Pre-processing

The pre-processing section provides the results from segmentation and feature extraction. From every pre-pressure and mass flow combination, 100 windows were generated from which ten features were extracted. Every window represents one data point with a total of 2700 data points for every fluid. This section will discuss the results of the feature extraction. This will be done while considering the fluidic parameter relations described in the theory section.

The first feature is the average voltage difference between the pressure sensors. The relation between fluid, mass flow and voltage difference is shown in Figure 13. The voltage difference is zero for a mass flow of zero. The voltage difference increases when increasing the mass flow. The voltage difference is directly related to the pressure and the mass flow is directly dependent on the pressure drop as it determines the force on the mass. This justifies the observation of an increased voltage difference when increasing the mass flow.

Furthermore, the steepness of the graph is related to the dynamic viscosity of the fluid which is observed as the difference in steepness for different fluids. The figure shows the voltage difference for six fluids. Isopropanol has the steepest curve while chloroform shows the least steep curve. In the theory section, a relation between pressure drop, mass flow and dynamic viscosity was proposed:

$$\frac{\Delta p}{Q_v} \propto \mu \tag{12}$$

A different slope between pressure difference and volume flow implies a different viscosity. Isopropanol has the highest viscosity of the six with 2.052 $mPa \cdot s$ and should thus create the steepest slope which is also observed in the figure. The slopes get less steep for fluids with a lower viscosity. However, chloroform has the smallest slope while hexane has the lowest viscosity. This observation can be ascribed to the difference in density between the fluids. The equation relates the volume flow to the viscosity. However, volume flow is not directly related to mass flow, instead the density should be taken into account. Higher densities will relate to more mass flow for lower volume flows. chloroform has more than double the density of hexane and thus has less volume flow for the same mass flow. The equation implies that a lower volume flow will relate to a lower pressure drop with the same viscosity. This explains why hexane has a steeper slope than chloroform.

The actuation frequency had a proposed inverse relation to the density of the fluid which is also observed in the data. The frequencies are shown in Figure 14. For example, chloroform is the most dense and has the lowest frequency while hexane has the lowest density and the highest frequency together with ethanol, isopropanol and methanol. The resolution of $100Hz$ does not allow the function to distinguish between these fluids while there is presumably a difference between the fluids. Moreover, water shows an actuation frequency of both $2400Hz$ and $2500Hz$ although this should not change. This is likely because the actuation frequency is close to the middle and gets rounded up or down due to limited resolution.

The second harmonic frequency should be located at approximately double the actuation frequency and should be in order of density. The distribution is shown in Figure 17. Hexane is $100Hz$ above double the actuation

frequency which implies that the actuation frequency of hexane must be between $2625 - 2650Hz$. The lowest frequency observed is from chloroform which is distributed between $4100 - 4200Hz$. The fluids are ordered exactly as expected with the frequencies increasing as the density of the fluid decreases.

The second harmonic frequency does contain more faulty data points with frequencies below the actuation frequency and above triple the actuation frequency. These frequencies are observed due to the range of the frequency analysis. The range was specified between $0 - 20KHz$ and the strongest magnitude is not necessarily the second harmonic as can be seen in Figure 15b. The magnitude of second harmonic is especially low for high pre-pressures. The magnitude spectrum is shown in Figure 10 which shows how strong the noise can be. This results in the findpeaks function returning the wrong frequency and thus causing wrong data points.

The magnitudes of both frequencies is dependent on the fluid as the magnitude is significantly lower for water than for the rest. Moreover, chloroform and hexane also show little deviation from ethanol, isopropanol and methanol. However, the relation is not fully understood and should be investigated.

The magnitudes of the frequencies are dependent on the pre-pressure but are not influenced by the mass flow as is shown in Figure 15. The pressure dependency is due to the stress exerted on the tube which causes the tube to slightly bend towards the flat top. This offset decreases the signal from the displacement sensors as the distance is greater and the capacitance change thus smaller.

The phase shift was calculated to determine the mass flow. The calculated phase shift for chloroform is shown in Figure 16. The figure shows the phase shift and standard deviation for different mass flows and pre-pressures. The phase shift shows the expected upwards trend when increasing the mass flow. However, not every mass flow corresponds to a unique phase shift. Therefore, the phase shift can only be used as an indication of the mass flow and not as a direct relation. Interesting is that the deviation decreases for higher pre-pressures, implying that the algorithm works better for these pressures.

There are two major influences to the inaccuracy of the phase shift. First, the phase shift can deviate over time and as every window is only a hundredth of a second, the observed phase shift in the window might deviate from the average phase shift. The second influence is the resolution of the actuation frequency. The phase shift is calculated by dividing the sample shift by the samples per period. The samples per period is calculated with the actuation frequency gained from the findpeaks algorithm. However, this actuation frequency has a resolution of $100Hz$. Therefore, the number of samples per period might deviate from the calculated samples per period. The phase shift is determined by dividing the lag over the calculated number of samples per period. Thus, the wrong actuation frequency can result in a wrong phase shift.

6.1.2 Machine Learning

The classification accuracy of the machine learning algorithms showed varying results under different implementations. This section will discuss the performance of these algorithms including an explanation on how they achieve this performance and the influence of the feature modification and hyper parameter optimization techniques.

The algorithms are provided with several features which hold relations to the physical quantities of the fluid. The density is directly related to the actuation frequency while the viscosity can be determined based on the voltage difference and phase shift as stated by the equation of Hagen-Poisseuille. The expectation is that the algorithms can use these relations to identify the fluids.

The decision tree was able to use these relations to classify the fluids with a maximum accuracy of 95%. The tree used the density to separate water, chloroform and hexane by splitting on both the actuation frequency and the second harmonic frequency. This gave a precision of 100% for these fluids. Furthermore, the viscosity was used for the classification of ethanol, methanol and isopropanol by splitting on the phase shift, related to the mass flow, and the voltage difference, related to pressure drop.

However, the decision tree cannot approximate Hagen-Poisseuille's equation and instead has to separate data points based on the features value. The tree classifies by nodes which separate data points. By looking at the voltage

difference shown in Figure 13, a splitting point is identified at 1.4×10^{-5} which separates isopropanol from all the other fluids. However, this kind of splitting cannot be performed for a mass flow of zero as the voltage difference is zero. This is the reason why approximately 10% of methanol, isopropanol and ethanol cannot be classified.

Normalization of the voltage difference which increases the accuracy of the decision tree from 76.08% to 95%. Normalization redistributes the voltage from a 10^{-6} to a distribution with mean zero and standard deviation of 1. The decision tree looks for splitting points in the data and should therefore not be influenced by normalization. However, the voltage drop is in a range too small for the algorithm to find splitting points. This was verified by testing the algorithm with the voltage drop multiplied by 10^6 resulting in the same accuracy for the normalized voltage difference.

The decision tree is able to attain the best accuracy when including the magnitudes into the feature set. The decision tree uses the magnitudes to split ethanol, methanol and isopropanol. However, these three fluids showed similar magnitudes and were thus expected to have the potential to identify the fluids. Therefore, further research needs to be performed to investigate the relation between the fluids and the observed magnitudes.

The performance of naïve Bayes is ascribed to the distribution of the frequencies and its inability to combine features. The decision of Naïve Bayes is based on the combined probability of every feature. Hexane, chloroform and ethanol have either a unique actuation or a second harmonic frequency. Therefore, the probability of one of these frequencies belonging to their respective class is 100%. However, ethanol, isopropanol and methanol need to be distinguished based on the combination of the voltage drop and phase shift. However, naïve Bayes is unable to combine the phase shift and voltage difference, instead the probabilities are used individually. The phase shift does not provide any information as it is equally distributed over the fluids. Moreover, the voltage difference provides little information. For example, a voltage drop of 2.5×10^{-5} has a probability of 100% to belong to isopropanol as seen in Figure 13. However, a voltage drop of 0.25×10^{-5} can belong to any of the other fluids, which explains the low precision and recall for ethanol, isopropanol and methanol. Moreover, this is the cause for the wrong classification of hexane and methanol after applying feature selection. Thus, the algorithm can distinguish fluids based on their density, but is unable to use the relation to viscosity.

The algorithm was able to attain the best accuracy by normalizing the voltage and including all the features except the pre-pressure and mass flow. The pre-pressure and mass flow are equally distributed over the fluids and thus have an equal probability to belong to any class. The performance of the algorithm increases when including normalization of the voltage difference. However, this is again, due to the range of the voltage which is verified in the same manner as for the decision tree. Lastly, the accuracy decreases when only selecting the actuation frequency, voltage difference and phase shift. This implies the algorithm was able to find the relations between the other features and the fluids such as the relation between the second harmonic frequency and hexane.

K nearest neighbour is able to use the relation between the physical quantities and the data to classify the fluids. The algorithm classifies by looking for the closest neighbours in the data space. The actuation frequency of chloroform and water are unique therefore, the neighbours on those axes are always of the same class. This makes the algorithm perform well for these classes. Furthermore, the algorithm is presumably able use the viscosity relation. The algorithm looks at every axis and can thus look at the phase shift and the voltage difference simultaneously. This combination leads to one specific point in the graph shown in Figure 13. The distribution of the voltage distribution is low enough for every point to have only neighbouring points of their own class. However, the relation between the mass flow and phase shift is inaccurate. Therefore, the optimal performance is not achieved. Nevertheless, the algorithm is able to combine the density and viscosity relations, making it an accurate classifier.

Normalization plays an important role for the performance of k nearest neighbour. It increases the accuracy from 62.47% to 92.96%. The algorithm looks for the neighbours with least absolute distance. However, before normalization, the features have different ranges and thus impact the absolute difference differently. For instance, the voltage difference is in the range of 10^{-5} while the actuation frequencies are $100Hz$ apart. After normalization, the features have similar ranges thus leading to similar distances between data points for every feature and thus improving the accuracy.

Hyperparameter optimization only marginally influences the performance of the classifier. The little influence of varying the k can be ascribed to the vast availability of data points. There are 2700 data points per fluid and 100 for every mass flow and pre-pressure combination. Furthermore, the neighbours are mostly distributed separately such as the actuation frequency and voltage drop. Therefore, ranging the number of neighbours from 1 to 100 does not influence the behaviour significantly. Not weighting the distance positively influences the accuracy of the algorithm. The cause is not fully understood however, it is likely because all features show some deviation. Therefore, data points from different classes can overlap and thus be located on the same point thus decreasing the accuracy when weighing the distance.

Overall, the algorithms performed well, but they relied on the discrete steps of the pressure drop. These discrete steps were due to the discrete steps for the mass flow. However, the algorithm should work for any mass flow. Therefore, a new training data set should be created with continuous mass flow data.

6.2 Improvements

The performance of the algorithms were limited by several aspects of the pre-processing including the segmentation and feature extraction phase. This section proposes improvements which should enhance the performance of the algorithms.

Segmentation can be improved by applying different filters and optimizing the windowing of the signal. This research applied filtering specifically for every feature. However, the signal contains high frequency components which do not provide useful information and should thus be filtered. The phase shift is observed by shifting the two signals across each other and determining the correlation. However, the high frequency components only distort the signals and therefore result in a lower correlation. Therefore, future research could potentially improve the accuracy of the phase shift by removing this noise.

Secondly, optimal window size and window overlap could be investigated. This research used 100 windows, chosen as a balance between data points per fluid and information per window. Moreover, no overlap was implemented. Future research could optimize the number of windows and introduce overlap to generate an optimal data set.

The main improvement for feature extraction in the frequency domain would be enhancing the resolution. Due to a window size of 2500 samples, only a resolution of 100 Hz was possible. The actuation frequency relates to the density of the fluid. Therefore, improving the resolution would enable the system to more easily distinguish between fluids. To improve the resolution, more samples need to be included in the signal. The actuation and second harmonic frequency are related to the density. This research was able to separate water, hexane and chloroform based on their actuation frequency. Increasing the resolution could result in a better accuracy. Therefore, either increasing the window size or up sampling the signal would prove beneficial for the accuracy of the machine learning algorithms.

Furthermore, faulty data acquisition could be prevented by only returning frequencies within the expected range. There are several data points which show wrong actuation frequencies and more, which extract the wrong second harmonic frequency. The actuation frequency has a minimal frequency of $2400Hz$ and the second harmonic $5200Hz$. Future research could improve the reliability of the function by changing the frequency range from $0 - 20KHz$ to approximately $2 - 5.5KHz$.

Enhancing the phase shift detection would improve the algorithms capability to approximate the mass flow. The phase shift detection could be improved by increasing the window size and by enhancing the resolution of the actuation frequency calculation. The mass flow is used to approximate Hagen-Poisseille's equation which is used to relate the features to the viscosity of the fluid.

Moreover, introducing a feature which combines the phase shift and voltage difference would presumably increase the accuracy of the algorithms. The combination of the two features relates to the viscosity of the fluid according to Hagen-Poiseuille's equation.

Finally, the models could be improved by using more steps between every mass flow, e.g. using steps of .5 g/h between every measurement. Currently, the models take advantage of the discrete steps in voltage difference observed when changing the mass flow. However, for a universal mass flow, the voltage difference will show a continuous spectrum. To ensure the models work regardless of the mass flow, the training data should take more steps between every mass flow.

6.3 Research potential

This research is the first step into combining artificial intelligence into the field of the mass flow sensors. These findings can be used for future research to improve the performance or using them for applications such as in the medical or industrial market. This section provides several implications that this research has on future research.

The algorithms could be trained to classify the fluids regardless of the temperature. The data generated by the IMFP sensor is dependent on the temperature. Therefore, the algorithm could be trained to learn this temperature dependence, by including several temperature measurements in the data set, ideally, learning the direct relation between data and temperature. This would result in a model which is able to classify fluids regardless of the temperature.

Regression algorithms could be used in future research for mass flow estimations and fluidic parameter estimation. By enhancing the accuracy of the phase shift detection, the mass flow could be detected. However, the algorithm could identify other dependencies between the mass flow and the data thus, creating a robust mass flow regression model without an accurate phase shift. Moreover, the viscosity and density of a fluid relates to the signals according to the relations described in the theory section. The linear relation observed in Figure 13 can presumably be estimated by linear regression. Therefore, these fluidic parameters could be estimated based on the actuation frequency, the phase shift and the observed voltage difference between the pressure sensors.

Moreover, a composition is identified by these fluidic parameters which thus can be used to estimate and classify the composition. Therefore, integrating regression has the potential to estimate the contributions of a fluid to a composition. [1]

7 Conclusion

The goal of this thesis was to create machine learning algorithms capable of classifying fluids using the signals from the integrated mass flow and pressure sensor. The available data consisted of four signals from the inlet and outlet resistive pressure sensors and the capacitive combs which measure the displacement of the tube. To apply machine learning, these signals needed to be converted into a data set. The algorithms used the data set to learn classification models which were later optimized using normalization, feature selection and hyperparameter optimization.

The signals from this sensor relate to the physical quantities of the fluids. Therefore, the signals can be used to identify the fluid. The observed relations were the relation between actuation frequency and density and between mass flow, pressure drop and viscosity. The feature extraction process focused on extracting this information from the data, this included: determining the frequency corresponding to the largest magnitude in the frequency domain of the displacement signals, determining the phase shift between the displacement signals and calculating the average voltage drop between pressure sensors.

The relation between the physical quantities and the features were accurate for the actuation frequency and the pressure drop but less so for the phase shift. The actuation frequency decreased for denser fluids. However, the resolution was too low to distinguish fluids with similar densities e.g. ethanol and methanol. The viscosity is related to the mass flow and pressure drop. This pressure drop was determined using the data from the resistive pressure sensors. Plotting the voltage difference against the mass flow for various fluids showed the linear relation. Lastly, the mass flow was extracted using the phase shift. However, the phase shift showed strong deviations between windows and did not accurately represent the mass flow. However, it could still be used in combination with the other features to identify the fluid.

Using this data, the decision tree, k nearest neighbour and naïve Bayes algorithms were able to generate models which could classify the fluids with an accuracy of 95%, 93% and 78.11%, respectively. The algorithms were able to identify the relations between the fluids and the features. However, the algorithms relied on the discrete steps in the mass flow sensor, therefore further research is required to enable classification using arbitrary mass flows.

Further research could be conducted into regression and implementing deep learning. The density and viscosity of fluids is related to the output of the sensor therefore, machine learning could potentially estimate these relations. Furthermore, deep learning has shown promising results in other field due to its ability to extract high-level features. The IMFP sensor provides various relations towards the fluid, some not even clearly understood. Therefore, deep learning could potentially use these relations to increase the performance.

In conclusion, the results from this research prove the potential of integrating artificial intelligence into the mass flow sensor. Using the other available methods such as regression and deep learning, new applications will surface such as composition estimation for drug administration. Thus, will prove useful for society.

References

[1] Dennis Alveringh. "Integrated throughflow mechanical microfluidic sensors". English. PhD thesis. Netherlands: University of Twente, Apr. 2018. ISBN: 978-90-365-4515-0. DOI: 10.3990/1.9789036545150.

[2] Dennis Alveringh, Remco J. Wiegerink, and Joost Conrad Lötters. "Integrated pressure sensing using capacitive Coriolis mass flow sensors". English. In: *Journal of microelectromechanical systems* 26.3 (June 2017), pp. 653–661. ISSN: 1057-7157. DOI: 10.1109/JMEMS.2017.2689162.

[3] Dennis Alveringh et al. "Improved capacitive detection method for Coriolis mass flow sensors enabling range/sensitivity tuning". Undefined. In: *Microelectronic engineering* 159 (June 2016). Micro/Nano Devices and Systems 2015, pp. 1–5. ISSN: 0167-9317. DOI: 10.1016/j.mee.2016.01.029.

[4] Dennis Alveringh et al. "Resistive pressure sensors integrated with a Coriolis mass flow sensor". English. In: *TRANSDUCERS 2017 - 19th International Conference on Solid-State Sensors, Actuators and Microsystems*. United States: IEEE, June 2017, pp. 1167–1170. ISBN: 978-1-5386-2733-4. DOI: 10.1109/TRANSDUCERS.2017.7994261.

[5] James Bergstra and Yoshua Bengio. *Random search for hyper-parameter optimization*. Feb. 2012. URL: https://dl.acm.org/doi/10.5555/2188385.2188395.

[6] Claesen et al. *Hyperparameter Search in Machine Learning*. Apr. 2015. URL: https://arxiv.org/abs/1502.02127.

[7] *findpeaks*. URL: https://nl.mathworks.com/help/signal/ref/findpeaks.html.

[8] *Learn By Implementation – K-Nearest Neighbor*. Sept. 2015. URL: https://depiesml.wordpress.com/2015/09/03/learn-by-implementation-k-nearest-neighbor/.

[9] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd. USA: Prentice Hall Press, 2009. ISBN: 0136042597.

[10] *U.S. Energy Information Administration - EIA - Independent Statistics and Analysis*. 2013. URL: https://www.eia.gov/todayinenergy/detail.php?id=7110.

[11] Raymond Veldhuis. *Lecture Notes Discrete-Time Signal Processing*. May 2019.

[12] Jindong Wang et al. *Deep Learning for Sensor-based Activity Recognition: A Survey*. Dec. 2017. URL: https://arxiv.org/abs/1707.03502.

8 Appendix

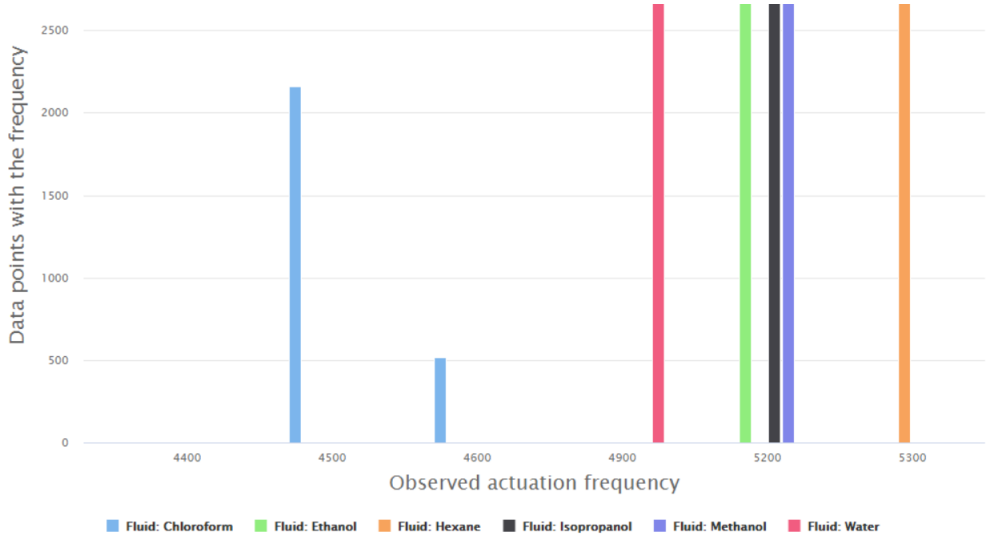


Figure 17: The number of example data per fluid with the specified second harmonic frequency.

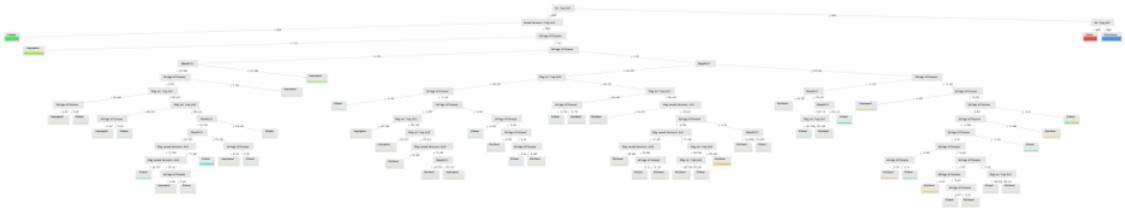


Figure 18: The complex DT generated after normalizing the voltage difference

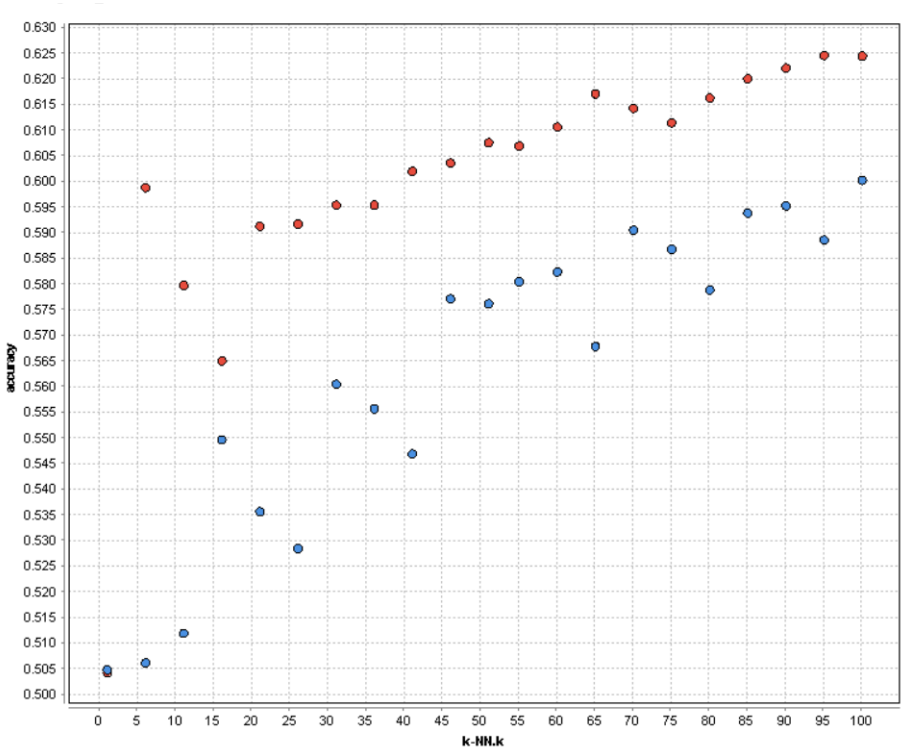


Figure 19: The accuracy of k-nn over different k's. The blue dots represent weighting the distance while the red dots do not.