İZMİR KÂTİP ÇELEBİ ÜNİVERSİTESİ
2010
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

Assessment of Fouling in Plate Heat Exchangers with Machine Learning Algorithms

Submitted to the Graduate School of Natural and Applied Sciences
in partial fulfillment of the requirements for the degree of

Master of Science

in Mechanical Engineering

by

Ceren Vatansever

ORCID 0000-0002-2287-9699

October 2022

This is to certify that we have read the thesis **Assessment of Fouling in Plate Heat Exchangers with Machine Learning Algorithms** submitted by **Ceren Vatansever**, and it has been judged to be successful, in scope and in quality, at the defense exam and accepted by our jury as a MASTER'S / DOCTORAL THESIS.

APPROVED BY:

Advisor: **Assoc. Prof. Dr. Ziya Haktan Karadeniz**
İzmir Kâtip Çelebi University

Committee Members:

Assoc. Prof. Dr. Sercan Acarer
İzmir Kâtip Çelebi University

Assoc. Prof. Dr. Alpaslan Turgut
Dokuz Eylül University

Date of Defense: October 6, 2022

Declaration of Authorship

I, **Ceren Vatansever**, declare that this thesis titled **Assessment of Fouling in Plate Heat Exchangers with Machine Learning Algorithms** and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for the Master's / Doctoral degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. This thesis is entirely my own work, with the exception of such quotations.
- I have acknowledged all major sources of assistance.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Date: 1.10.2022

Assessment of Fouling in Plate Heat Exchangers with Machine Learning Algorithms

Abstract

Fouling is the accumulation of undesired particles on heat transfer surfaces which affects the heat transfer performance of a heat exchanger negatively. The accumulation of these particles prevents heat from being transformed through the heat exchangers by generating a fouling layer-like insulation. The main aim of the thesis is to investigate the machine learning algorithms to classify and predict the fouling status of PHE used in combi-boilers, to generate the background of the predictive maintenance, besides investigating the fouling effect on PHEs in terms of heat transfer and energy consumption by using a 1-D model.

The required data to train the machine learning algorithms is acquired experimentally by using an artificially generated method for evaluating the fouling behavior. The effect of fouling on PHE performance is assumed as similar to the performance loss that would be occurred if the PHE that is already used in the combi-boiler, would be replaced with a PHE that has fewer plate numbers. The experiment results show that the expected trends of output temperatures and pressure drop values of both channels are seen.

The overall heat transfer coefficient and fouling resistance coefficient are calculated as the performance values of the tested PHEs. As expected, the overall heat transfer coefficients are resulted in decreasing while the fouling resistance coefficient is increasing.

The 1-D numerical model is generated by using Runge Kutta 4^{th} order ordinary differential equation solving method. The differential equations are created based thermal resistance method for both channels to evaluate the temperature distributions by using the experimental data. The results show that with less than 2% error the model is concluded to receive the correct outputs with the experimental outputs.

The additional required power to reach the setpoint of DHW defined by the customer is calculated at maximum fouling by using the model results. The results show that combi-boiler appliances need to supply approximately 16 and 7 kW additional heat output to reach the required setpoint of DHW in case of maximum fouling for 32 and 30 plates PHE, respectively.

The obtained data is implied to train the machine learning algorithms, Naïve Bayes, k-nearest neighbor, and decision tree. The k-fold cross-validation method is used to avoid overfitting for the implementation method. It results that the k-nearest neighbors model would be the best among the other models for predicting the classes according to the overall heat transfer coefficient values. The decision tree model results show that the model is independent of its maximum number of splits selection. The results show the decision tree model gives better performance in classifying than the Naïve Bayes model according to the accuracy results.

Keywords: Fouling, machine learning, plate heat exchangers, classification, 1-D modeling, Runge Kutta, combi-boiler

Plakalı Isı Değiştiricilerde Kirliliğin Makine Öğrenmesi Algoritmaları ile İncelenmesi

Öz

Kirlilik, bir ısı değiştiricinin ısı transfer performansını olumsuz yönde etkileyen, ısı transfer yüzeylerinde istenmeyen parçacıkların birikmesidir. Bu parçacıkların birikmesi, yalıtım benzeri bir kirlilik tabakası oluşturarak ısının ısı değiştirici aracılığıyla aktarılmasını engeller. Tezin temel amacı, kombilerde kullanılan plakalı ısı değiştiricilerin kirlenme durumunu sınıflandırmak ve tahmin etmek için makine öğrenmesi algoritmalarını araştırmak, kestirimci bakımın arka planını oluşturmak, ayrıca plakalı ısı değiştiriciler üzerindeki kirlenme etkisini ısı transferi ve enerji açısından 1-B model kullanarak incelemektir.

Makine öğrenimi algoritmalarını eğitmek için gerekli veriler, kirlenme davranışını değerlendirmek için yapay olarak oluşturulmuş bir yöntem kullanılarak deneysel olarak elde edilir. Kirlenmenin plakalı ısı değiştirici performansı üzerindeki etkisi, kombide halihazırda kullanılan plakalı ısı değiştiricinin daha az plaka numarasına sahip bir plakalı ısı değiştirici ile değiştirilmesi durumunda oluşacak performans kaybına benzer olarak kabul edilir. Deney sonuçları, her iki kanalın çıkış sıcaklıklarının ve basınç düşüş değerlerinin beklenen eğilimlerinin görüldüğünü göstermektedir.

Toplam ısı transfer katsayısı ve kirlenme direnci katsayısı, test edilen plakalı ısı değiştiricilerin performans değerleri olarak hesaplanır. Beklendiği gibi, toplam ısı transfer katsayıları, kirlenme direnci katsayısı artarken azalmaktadır.

1-B sayısal model, 4. Mertebeden Runge Kutta diferansiyel denklem çözme yöntemi kullanılarak üretilmiştir. Diferansiyel denklemler, deney verilerini kullanarak sıcaklık dağılımlarını değerlendirmek için her iki kanal için de termal direnç yöntemi temel alınarak oluşturulmuştur. Sonuçlar, %2'ten daha az hata ile modelin, deney çıktılarıyla doğru çıktıların alındığı sonucuna varıldığını göstermektedir.

Müşteri tarafından tanımlanan DHW set noktasına ulaşmak için gereken ek güç, model sonuçları kullanılarak maksimum kirlenme için hesaplanmıştır. Sonuçlar, maksimum kirlenme durumunda gerekli DHW set noktasına ulaşmak için kombi cihazlarının sırasıyla 32 ve 30 plakalı ısı değiştiriciler için yaklaşık 16 ve 7 kW ek ısı çıkışı sağlaması gerektiğini göstermektedir.

Elde edilen veriler, makine öğrenmesi algoritmaları, Naive Bayes, k-en yakın komşu ve karar ağacını eğitmeye yöneliktir. Uygulama yöntemi olarak modelin fazla uydurmasını önlemek için çapraz doğrulama yöntemi kullanılır. Sınıfları toplam ısı transfer katsayısı değerlerine göre tahmin etmek için k-en yakın komşu modelinin diğer modeller arasında en iyisi olacağı sonucuna varılmıştır. Karar ağacı modeli sonuçları, modelin maksimum ayrım değeri seçiminden bağımsız olduğunu göstermektedir. Sonuçlar, karar ağacı modelinin doğruluk sonuçlarına göre Naive Bayes modeline göre sınıflandırmada daha iyi performans verdiğini göstermektedir.

Anahtar Kelimeler: Kirlenme, makine öğrenmesi, plakalı ısı değiştiriciler, sınıflandırma, 1-B modelleme, Runge Kutta, kombi

Acknowledgment

I am deeply expressing my sincere gratitude for the help that is provided to me during the whole study process by my consultant, Assoc. Prof. Ziya Haktan Karadeniz. I would like to express my sincere thanks and respect to the Bosch Thermotechnology R&D family, my colleagues, Ahmet Kuzucanlı and Alp Emre Yaşar, and my section manager in the hydraulic module department Erden Akdarı who made this study possible, and pioneered make this work happen.

Furthermore, I would like to express my genuine gratitude to my family, friends, and my unconditional supporter Ali Haydar Özay who has always been by my side at every moment of this study.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Al_2O_3	Aluminum oxide
CFD	Computational fluid dynamics
CH	Central heating
DCW	Domestic cold water
DHW	Domestic hot water
FTIR	Fourier Transform Infrared Spectrophotometer
HVAC	Heating Ventilation and Air Conditioning
kNN	k-Nearest Neighbor
LMTD	Logarithmic mean temperature difference
PHE	Plate heat exchanger

List of Symbols

μ_f	Dynamic viscosity of the fluid [Pas]
μ_w	Dynamic viscosity of the fluid at wall temperature [Pas]
ρ	Density [kg/m^3]
ΔT_{lm}	Logarithmic mean temperature difference
A	Area [m^2]
c_p	Specific heat at constant pressure [J/kg°C]
d_i	Euclidean distance
D_h	Hydraulic diameter [m]
h	Convection heat transfer coefficient [W/m^{2}K]
$\dot{m}$	Mass flow rate [kg/s]
k	Conduction heat transfer coefficient [W/mK]
K	Number of classes
Nu	Nusselt number
p	Number of predictors
p	Wetted perimeter [m]
P	Posterior probability
Pr	Prandtl number
$\dot{Q}_c$	Heat transfer rate of cold channel [J/s]
$\dot{Q}_h$	Heat transfer rate of hot channel [J/s]
$\dot{Q}_{total}$	Total heat transfer rate [J/s]
R_{CH}	Thermal resistance of CH channel
R_{DHW}	Thermal resistance of DHW channel

R_f	Fouling resistance coefficient
Re	Reynolds number
T	Temperature [°C]
U	Overall heat transfer coefficient [W/m^{2}K]
V	Velocity [m/s]
x	Predictors
y	Responses
z	Training and testing data set

Chapter 1

Introduction

1.1 Fouling in Plate Heat Exchangers

Fouling is defined as the process of accumulation of undesired particulates on heat transfer surfaces [1]. The accumulation of these particulates causes a lack of heat transfer in the heat exchangers. The plate heat exchangers mostly used water as a fluid, thus the fouling problem is often occurred by the accumulation of particles coming from the other components of old, rusty installments or by precipitation of particles that contain a high amounts of calcium compounds. If the fouling is occurred by the accumulation of undesired particles, especially rusty metal compounds, the fouling type is named as particulate fouling. If the fouling is occurred by precipitation of calcium compounds contained in water, this type of fouling is named precipitation fouling. In addition, if the surface of the heat exchanger is exposed to corrosion of itself by water, particle dissociation from the surface and particle aggregation on the surface may have occurred. This type of fouling is named as corrosion fouling. These fouling types can be seen together as individually. If there is more than one fouling type is affected on the heating surface, the fouling type is named composite fouling. Especially, particulate fouling and precipitation fouling occurred together.

Combi-boiler is a heating appliance that is used to heat the residence and provide domestic hot water to the customer. Domestic hot water is denominated for the hot water used in the shower/bath or the kitchen. The domestic hot water, whether provided via tank or instantaneous heating of mains water, is heated in the combi-boiler appliances by using plate heat exchangers (PHE). The particulate fouling is encountered in the PHEs, due to the undesired rusty particles coming from the installments. The precipitation fouling is encountered in the PHEs, because the calcium compounds lose their solubility in the water when the temperature of the water is increased, i.e., the water is heated.

1.1.1 Combi-boilers

A combi-boiler contains three main parts, the outside structure, the hydraulic part, and the heat cell part. The heat cell part is where the combustion of gaseous fuel, e.g., natural gas, is occurred. In Figure 1.1, the heat cell is designated as 1. The heat that is meant to transfer to the water is generated in this part of the combi-boiler. This heat transfer has occurred via a fuel-water heat exchanger, which is called a primary heat exchanger in combi-boilers. The water that is heated in primary heat exchangers has two functions. One is circulating through the radiator’s line and heating the residence, the other function is circulates through the second closed loop, through the PHE to heat the domestic hot water.

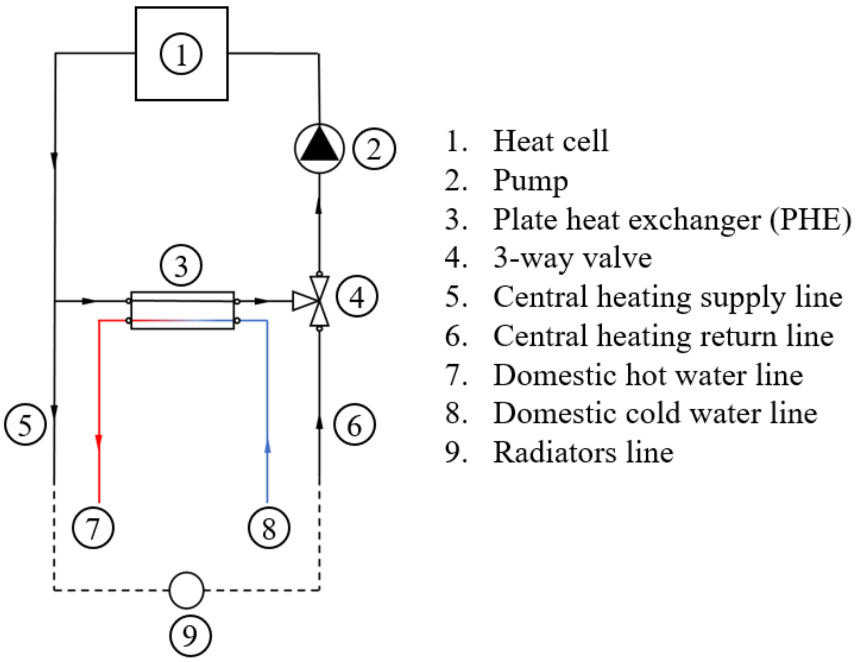


Figure 1.1: The schematic of the combi-boiler

The other essential part of the combi-boiler is the hydraulic. The water circulating pipes, the PHE, and the main controlled sensors are placed in this part. The PHE has two immiscible channels, one channel is for the water heated by the heat cell and goes through the line of radiators, central heating water (CH), the other channel is for the domestic hot water which is coming as mains water and goes to the household for domestic usage, domestic hot water (DHW). In Figure 1.1, the PHE and its inlets and outlets are designated as 3.

The CH water is circulated via a pump, as shown in Figure 1.1 as 2. The diverting of the water to the radiators line or PHE is provided by a 3-way diverter valve, shown in Figure 1.1 as 4. The diverter valve is controlled by the combi-boiler's control unit. When the customer opens the tap for hot water, in the DHW line there is flow that is sensed by a flow sensor and transferred to the control unit. Then, the control unit sends a signal to the diverter valve to change its shaft to divert the water to the DHW line from the CH line.

1.1.2 Fouling Behaviour

Plate heat exchangers, which are used to transfer heat indirectly from the combustion gases to the domestic water side, are constantly under the influence of small particles and water-soluble compounds coming from other components of the heating system (combi-boiler parts, pipes, radiators, etc.) and the mains water. When the accumulations that cause blockage in the heat exchangers are examined, it is seen that the particles detached during the flow from the corrosion layer on the inner surfaces of the system components and the precipitation of the calcium compounds contained in the mains water on the surfaces with the effect of temperature change are the main factors causing the blockage.

It is known that accumulations occur differently on the cold (mains water) and hot (central heating) flow sides of the plate heat exchanger. The accumulation on the mains water side occurs only because of the precipitation of crystals in the water. There are many studies in the literature examining this subject [2-6].

In these studies, the effect of water physio-chemical properties (temperature, pH, concentration, etc.) and heat exchanger geometry on the precipitation amount was investigated [2, 3, 6].

In addition, the effect of the precipitation of calcium compounds on the mains water side of the plate heat exchanger on the performance of the plate heat exchanger has been studied in detail [7].

On the central heating side of the plate heat exchanger, there is a closed loop between the radiators and the boiler. Since there is no continuous mains water supply in the central heating cycle, compared to the domestic water side, the central heating side of the plate heat exchangers is much less affected by the pollution caused by the precipitation of calcium compounds [7].

Since the main material of the primary heat exchanger is aluminum, corrosion occurs on the surfaces in contact with water. In the examination carried out by Bosch Thermotechnology on used plate heat exchangers in this study, particle accumulation was observed in the plate heat exchanger. In addition, the particles causing clogging on the heating water side were investigated by FTIR (Fourier transformation infrared) analysis. As a result of the analysis, it was concluded that the particles that cause blockage because of high amounts of aluminum and oxygen element and low amounts of calcium element originate from the primary heat exchanger. The reason why the calcium element is found as a result of the analysis is that although it is a closed cycle without an active water supply, there are calcium crystals ($CaCO_3$) crystals in the water in the CH line. It is known that the solubility of this calcium compound in water decreases at high temperatures. For this reason, the presence of calcium, which also allows the Al_2O_3 particles to be attached, affects particle accumulation. This phenomenon is also studied in the study by Zhang et. al [6]. They studied the Al_2O_3 particulate accumulation alone and then together with the calcium compounds. The result that is obtained that the fouling resistance coefficients which indicate the fouling layer thickness, are higher than the Al_2O_3 particulate accumulation studied alone.

There are many factors that determine the cross section narrowing and the amount of pollution in the plate heat exchanger. The main of these factors are:

1. The source, type and amount of the particles causing the blockage,
2. Plate heat exchanger geometry,
3. Flow characteristics (fluid velocity).
4. Usage time

With the FTIR analysis performed by Bosch Thermotechnology, it was seen that the biggest cause of blockage in the plate heat exchanger was the Al_2O_3 (alumina) compound.

In addition, the information coming from the field shows that in houses where the highest congestion failure is encountered, the old installations used show that other system components other than the combustion chamber also cause particle accumulation.

Assuming that the installations cannot be changed, this source of pollution will not be considered as a parameter.

The primary heat exchanger in the heat cell is produced by casting method from aluminum material in a cylindrical structure that will allow the fuel to burn in it.

There are water passage channels that follow a helical path on the outer wall. There are contact of water with the aluminum surface over a large surface area. The parameters that affect the corrosion of aluminum with water are given below.

1. Temperature
2. Flow Rate
3. pH

When aluminum reacts with water, a natural oxide layer is formed on the wet surface, which provides resistance to the corrosion of the aluminum with mass loss. This oxide film layer can be examined in two layers the inner layer in direct contact with the metal and the outer layer in contact with water, whose structure changes depending on the temperature. The outer layer is active in electrochemical reactions with water due to flow and changing temperatures. The inner layer that encounters the metal continues to form as a result of the reactions with the aluminum. The corrosion of aluminum depends on the thickness of the oxide layer formed. The thickness of the oxide layer formed on the surface increases with temperature and time. As the temperature rises, the solubility of soluble gases (especially oxygen) in water decreases. This means more oxygen is available to react with aluminum [8].

1.1.2.1 Temperature Effect

Up to 60-70 °C water temperature, the oxide film, which is a thin layer, cannot show sufficient resistance to the reactions with the ions in the water, and pitting corrosion is dominant on the aluminum surface. It is seen that the pitting depth at the surface decreases as the temperature rises (Figure 1.2). At temperatures of 70 °C and above, the tendency to pitting corrosion gradually disappears. This is because the oxide layer tends to thicken and fill the pits.

Table 1.1: Corrosion forms according to temperature change [8]

Temperature	Corrosion Forms
<100 °C	Pitting corrosion (above 60-70 °C in tap water, pitting corrosion tends to decrease)
100 –150 °C	Uniform corrosion
150 – 250 °C	Uniform corrosion and intergranular corrosion
>250 °C	Intergranular corrosion (with metal destruction)

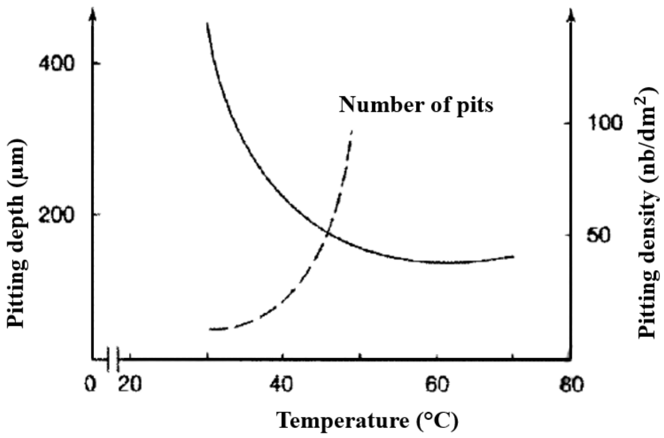


Figure 1.2: The pitting corrosion behavior according to temperature change [8]

Between 100 and 150 °C, uniform corrosion is observed, which causes reductions in the cross-sectional area with wear affecting the entire surface area. At 150 °C and above, the effect of intergranular corrosion begins to be seen. The relationship of temperature with corrosion is summarized in Table 1 [8].

In simulations made by Bosch Thermotechnology, the temperature distribution on the aluminum wet surface of the corroded primary heat exchanger was investigated. The water temperatures during the operation were taken as the design temperatures of 60 and 80 °C, respectively, at the inlet and outlet of the primary heat exchanger.

1.1.2.2 Flow Rate Effect

In a situation where all other parameters are kept constant, the corrosion effect of stagnant water on the wet aluminum surface is greater than that of water flowing up to a certain speed.

At high velocities, the effect of the flow is seen as erosion on the aluminum wet surface. Aluminum can withstand erosion effects up to 2.5 – 3 m/s [8]. The highest average speed seen in the primary heat exchanger in existing Bosch combi devices is 1.5 m/s. When viewed from a general point of view at these flow rates, it is expected that there will be no erosion effect on the surface. It was determined that the variable cross-sectional area of the water passage channel in the primary heat exchanger and the Reynolds number varied between 20795 and 59297. Assuming the critical Reynolds number of 2300 for in-channel flow, the flow in the primary heat exchanger is determined to be turbulent. For this reason, it is possible to see instantaneous and locally high velocities and therefore erosion of wet surfaces.

The oxide layer formed on the aluminum surface is not stable in flowing environments. At low velocities, there is a decrease in reaction rates relative to the stagnant flow. Reduction in reaction rates shows a decrease in mass loss from the wet surface in the case of pitting corrosion (at temperatures below 60 °C), and a decrease in the rate of layer formation in the formation of an oxide layer.

At high flow rates, aluminum becomes vulnerable as the flow breaks the oxide layer formed from the surface. As the aluminum becomes unprotected, the corrosion effect on the surface increases and mass loss occurs. The cycle between the pump not working and working conditions in the boiler can create an unstable and unpredictable cycle of accumulation and rupture on the surface. Therefore, the rupture of the corrosion layer in the primary heat exchanger in the combustion chamber depending on the flow rate is another parameter that causes accumulation in the plate heat exchanger.

1.1.2.3 pH Effect

While the pH value of water is between 5 and 8, the solubility of aluminum oxide in water is much less than in more acidic or basic environments (Figure 1.3).

In Figure 1.3, solubility is expressed with the concept of normality, which is the equivalent number of grams of the substance dissolved in one liter of solution. In acidic and basic environments where aluminum oxide dissolves, the aluminum surface remains unprotected. Therefore, corrosion, which causes particle breakage to cause mass loss, is more common in acidic and basic environments [9, 10].

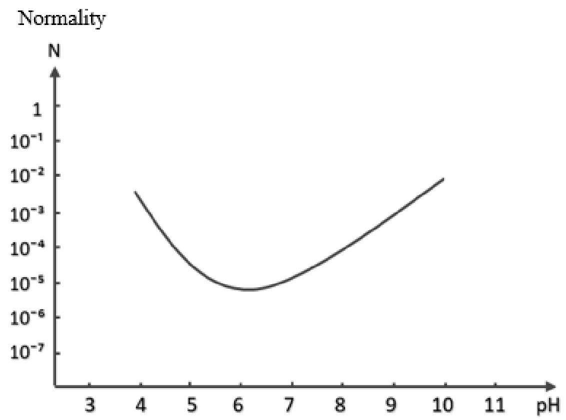


Figure 1.3: The variation of aluminum oxide solubility in water with pH [8]

In the reactions that occur because of the contact of aluminum with water, H+ and OH- ions are formed which can change the pH of the environment [10]. In the pits formed because of pitting corrosion, where these reactions are intense, the pH value of the water changes to acidic and basic [11].

It is predicted that the reaction rate changes with more than one parameter such as temperature, flow rate, according to the operating dynamics of the boiler. It is foreseen that this reaction rate change causes small changes in the pH value of the water, and that because of this pH change, the reactions may accelerate.

Generally used tap water has a pH between 6.5 and 8.5. Although partial pH changes can be seen, the pH value of water in general is the range in which aluminum oxide shows low dissolution. Therefore, the effect of pH value on corrosion will not be considered in this study.

1.2 Predictive Maintenance

At the beginning of the fouling process, the accumulations only affect the flow locally, grow as time progresses and become in a position to affect the efficiency of the plate heat exchanger completely, such as narrowing and congestion in the cross sections of channels. Thus, the prevention of fouling is essential to avoid the lack of customer comfort.

The prevention of fouling is recently associated with predictive maintenance. In recent time, the well-known maintenance process is reacting to the problem of a machine at the time the failure occurs. If maintenance has occurred in a schedule that the time is decided according to statistics not real-time data, this maintenance type is only preventing. During this preventing maintenance, there can be a time that the machine can work without failure, however, it is not known. This causes unnecessary maintenance costs. If the need for maintenance of a machine would be known, the maintenance would be carried out just in time and need. The predictive maintenance concept contains this phenomenon. If the status of a machine is monitored and processed during the machine’s operating hours, and the failure point can be trained to an algorithm, the maintenance time can be predicted. The status classification of the machine can be obtained by using machine learning techniques.

In this thesis, the fouling status, i.e., failure status caused by fouling, would be known by applying the machine learning algorithms for classification. For the given data, models are trained to classify the unseen data that would be encountered during the operation of the combi-boiler.

In the literature, modeling and prediction algorithms have become popular study subjects recently in the area of fouling prevention or predictive maintenance. These algorithms, which include algorithms for prediction and detection based on autoregressive integrated moving average (ARIMA) [12], auto-associative kernel regression (AAKR) [13], support vector machines (SVM) [14, 15], and artificial neural networks (ANN) [16-20], have largely been based on statistical methods and machine learning algorithms. Kalman filter research has also looked at model-based fouling prediction [21].

An algorithm to forecast fouling behavior has also been examined with the predictive maintenance technique. The design of the predictive maintenance procedures is mostly based on data-driven fault diagnosis. This algorithm's goal is to identify abnormalities, and fault diagnostics often concentrate on statistical techniques that offer classification and clustering. The majority of failure mechanisms are connected to deterioration processes [22].

By keeping an eye on the health system, as in reference, the data collection procedure may be maintained [23]. Naive Bayes, k-nearest neighbors (kNN), decision trees, and random forests are some of the machine learning methods used for classification. These algorithms are successfully researched by Shohet et al. [24] to classify boiler defects using simulation results. The decision tree model provides the best outcome with 97.8% accuracy as a result. As observed from the references, machine learning algorithms are typically utilized in the HVAC business, particularly in heat exchangers, but when open resources are taken into account, the classification machine learning algorithms on PHEs have barely been examined.

Chapter 2

Method

The process of the study is carried out in titles of data acquisition, data implementation of partial blockage assessment and data implementation of fouling assessment methods.

2.1 Data Acquisition

A data acquisition method is generated to simulate the fouling behavior in PHEs. The data that is needed to imply in algorithms, is acquired by using experimental methods. The experiment process is maintained with experiment procedure design, experiment parameters determination, experiment test rig setup and data reduction, respectively.

2.1.1 Experiment Procedure Design

Healthy and faulty data is needed to train the algorithm which is required for the algorithm to distinguish. Healthy data stands for the zero-hour performance of PHE while there is no fouling and PHE has its most effective status, i.e., is healthy.

The PHEs are designed to be used in combi-boilers according to the heat transfer requirements of combi-boilers. The power outputs of combi-boilers, i.e., heat transfer outputs, in product portfolios are generated by companies regarding the general needs of households. The PHEs are designed to meet this required power output and carry out the required heat transfer from heated water by natural gas to DHW in determining volume flow rates.

The volume flow rate of the DHW line is limited and controlled with a flow limiter device in combi-boiler utilities to keep the volume flow rate constant. The DHW goes through one channel of PHE. The other channel, the CH water line is circulated with a pump. When DHW usage is required by the customer, the combi-boiler operates at its maximum power, thus the volume flow rate of CH water is kept constant. Therefore, PHEs are designed to maintain a temperature difference with the specific two flow rates of its two channels to provide the required heat transfer.

According to the mentioned requirements, the number of plates of PHEs is determined to be used in combi-boilers which gives specific maximum power. Each PHE with a number of plates has a technical specification that indicates the temperature difference corresponding to volume flow rates of both channels. These data that are generated in design process of PHEs, are stood for the zero-hour performance, when there is no fouling in channels. Consequently, these technical design data of PHEs for particular number of plates are used as healthy data that denotes the reference performance values to imply algorithms.

Faulty data stands for the performance data (outlet temperatures of channels, pressure drop of channels) of PHEs after fouling starts. The faulty data is generated from experiments.

During the design process of the experiment procedure, it is assumed that when fouling in PHE starts, the effect of fouling on the performance of PHE would be the same if the PHE has a smaller number of plates would be used instead of the designed one according to the combi-boiler power output. Therefore, to simulate the fouling behavior in PHEs, the technical specifications of a PHE are applied as experiment parameter to a PHE that has a smaller number of plates.

2.1.2 Experiment Parameters

The volume flow rates of the DHW line and CH water line are kept as constant as it is possible during operation in real life of a combi-boiler. Therefore, the volume flow rates of both water lines are kept constant during tests. The used volume flow rate values in liter per minute (l/min) are shown in Table 2.1.

The mentioned technical specifications of PHEs that have 30 and 32 plates are applied as test conditions to PHEs that have 28, 26, 24, 22, 20, 18 and 16 plates. The DHW and CH volume flow rate values in the technical specification of a PHE that has 32 plates, are used as a test condition that would give the reference performance values of a PHE with 32 plates, indicating the healthy value, as shown in Test 1. Similarly, the used volume flow rates of both channels for a PHE that has 30 plates as a test condition are the volume flow rate values in the technical specification of a PHE with 30 plates, also it would give the healthy value, as shown in Test 10.

Table 2.1: Experiments and conditions applied to demonstrate the clogging behavior of the PHEs

	Condition 1				Condition 2		
	Tested PHE plate number	CH Flow Rate (l/min)	DHW Flow Rate (l/min)		Tested PHE plate number	CH Flow Rate (l/min)	DHW Flow Rate (l/min)
Test 1	32	29	18	Test 10	30	26	10.3
Test 2	30	29	18	Test 11	28	26	10.3
Test 3	28	29	18	Test 12	26	26	10.3
Test 4	26	29	18	Test 13	24	26	10.3
Test 5	24	29	18	Test 14	22	26	10.3
Test 6	22	29	18	Test 15	20	26	10.3
Test 7	20	29	18	Test 16	18	26	10.3
Test 8	18	29	18	Test 17	16	26	10.3
Test 9	16	29	18				

Furthermore, except Test 1 and 10 the rest of the Tests shown in Table 2.1 represent the tests that would give results the faulty data. In the fault tests (Test 2,3,4,5,6,7,8,9,11,12,13,14,15,16 and 17), the volume flow rates from technical specifications of 32 and 30 plates are used for the PHEs that have 28 to 16 plate numbers, as it is shown in Table 2.1. So as in Test 2, when the 32 plates PHE technical specifications are applied to 30 plates PHE, the result is assumed to show a performance decrease if the 2 plates of 32 plates PHE are clogged. Similarly, the same inference can be deduced for other test conditions. In the end, in Test 9, clogging of 16 plates, 50% clogging for 32 plates, is evaluated as the worst case.

The DHW and CH inlet temperatures are kept constant for all test conditions, 1 to 17. 10^{0}_{-1} °C and 72^{1}_{-1}°C are DCW and CH water line inlet temperatures, respectively. The inlet temperatures are taken from the design inlet temperature parameters for PHEs according to the inlet temperatures that the PHE would be most exposed to during the operation of the combi-boiler in real life.

2.1.3 Test Rig

The PHEs that have 32 and 30 plates are tested stand-alone in the test rig according to the test conditions shown in Table 2.1. The stand-alone test demonstrates a test in which only the component, i.e., PHE is tested with the conditions that would be occurred in real life in combi-boiler. The inlet temperatures and flow rates of both channels of a PHE are used as in 2.1.1.

The stand-alone test rig simulates the DHW line and CH line as in real life as possible as it is. The experimental setup contains two lines that represent CH line (orange colored) and DHW line (green colored) circuits (Figure 2.1).

In both lines, pneumatic valves are placed to direct the water to the required line. Also, both lines have flow control valves and flow meters to provide information to control and measure the volume flow rate of water circuits. The DHW line simulates the open flow circuit as in real life. The CH water line simulates the closed loop circuit as in real life in combi-boilers. The CH line has a pump to circulate the water through the closed loop. A tank is used to store the heated water by another closed water loop which is heated by a combi-boiler. Here the combi-boiler is used as a heating source only. The yellow-colored line represents the gas line that the combi-boiler needs during operation. A gas valve is placed on the gas line to control the gas passing through the combi-boiler for safety reasons.

The PHE is shown as tested PHE in schematic. Pressure difference and inlet and outlet temperatures are measured from the temperature and pressure differentiation sensors that is placed in both lines.

There is another plate heat exchanger is used in the test rig to control the CH inlet temperature to test PHE. The simulated CH water line is heated by a combi-boiler although non-directly. There is no combi-boiler automatic control to adjust the CH inlet temperature to test PHE in the rig. Therefore, a heat exchanger is placed to provide cooling to the heated CH water line to adjust the required inlet temperature of CH to test PHE.

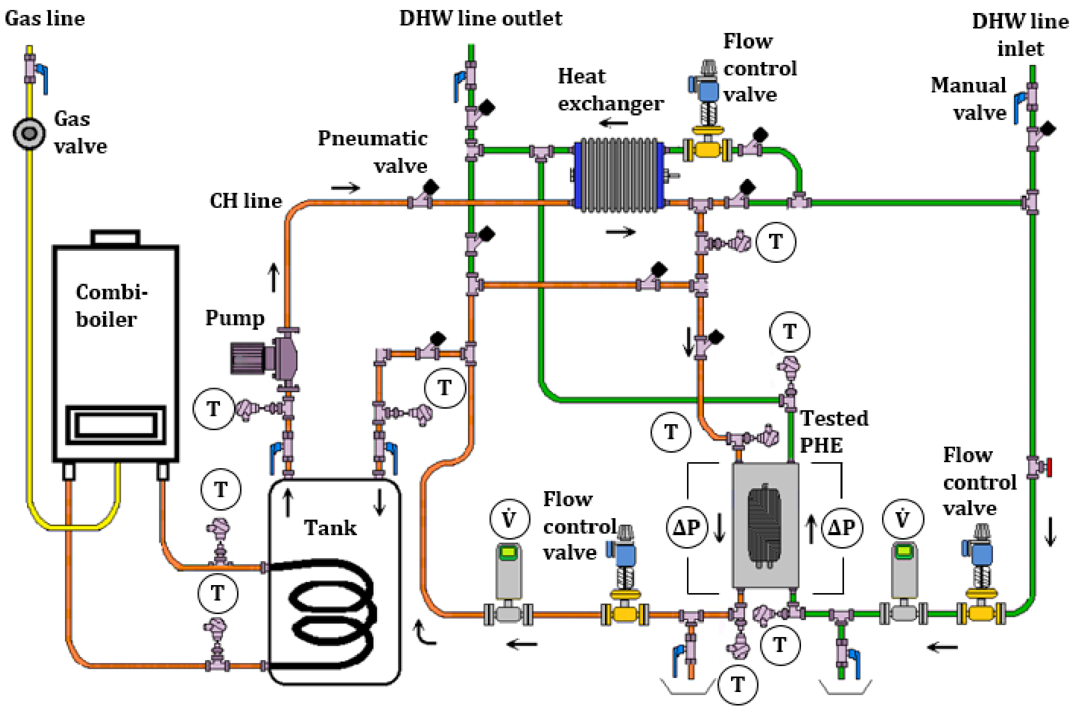


Figure 2.1: Stand-alone test rig schematic for PHE

During tests, the steps listed below are followed.

1. Place the PHE to be tested.
2. Supply water to CH closed loop up to 2 bar statistic pressure.
3. Turn on the combi-boiler and adjust the set temperature according to the required one in CH closed loop.
4. Turn on the CH circuit pump and adjust the modulation percentage according to get the required volume flow rate.
5. Use a flow control valve to get better accuracy in volume flow rate.
6. Heat the CH water to the required temperature.
7. Use the additional heat exchanger by providing cooling to achieve better accuracy in temperature.
8. Adjust the set point of the chiller system.
9. Turn on the DCW valve in the chiller line to supply the DCW to the PHE.

10. Adjust the flow control valve placed in the DCW line to get the required volume flow rate.
11. Measure the outlet temperatures of PHE channels and pressure difference.
12. Repeat the steps for the new PHE to be tested.

2.1.4 Data Reduction

The most essential effect of fouling on PHE is performance decreasing, i.e., pressure difference increasing and heat transfer efficiency decreases.

The obtained data from experiments are inlet and outlet temperatures, volume flow rates and pressure differences of CH and DHW channels of PHE. The overall heat transfer coefficient and fouling resistance coefficient are calculated from these data that are obtained from experiments to evaluate the performance behavior of PHEs.

The overall heat transfer coefficient is calculated by using the logarithmic mean temperature difference (LMTD) method. In Equation (2.1), the total heat transfer rate should be calculated first to get the overall heat transfer coefficient that is denoted as U. The area, which is denoted as A, is taken as the heat transfer area, i.e., the projection area of plates. LMTD is calculated as in Equation (2.2) for every test condition.

$$\dot{\mathrm{Q}}_{\mathrm{total}} = \mathrm{UA\Delta T_{lm}} \tag{2.1}$$

$$\Delta\mathrm{T_{lm}} = \frac{\Delta\mathrm{T_1} - \Delta\mathrm{T_2}}{\ln\left(\Delta\mathrm{T_1}/\Delta\mathrm{T_2}\right)} \tag{2.2}$$

Here the ΔT_1 and ΔT_2 are representing the temperature difference between two channels at the inlet and outlets of the PHE. Due to counter-flow in PHE, ΔT_1 and ΔT_2 are calculated as in Equation (2.3) and (2.4).

$$\Delta T_1 = T_{h,in} - T_{c,out} \tag{2.3}$$

$$\Delta T_2 = T_{h,out} - T_{c,in} \tag{2.4}$$

$T_{h,in}$ and $T_{h,out}$ denote the inlet and outlet temperatures of hot channel, i.e., CH water channel. $T_{c,in}$ and $T_{c,out}$ denote the inlet and outlet temperatures of cold channel, i.e., DHW channel. The inlet and outlet temperatures of both channels and the temperature differences (ΔT_1 and ΔT_2) are represented in Figure 2.2.

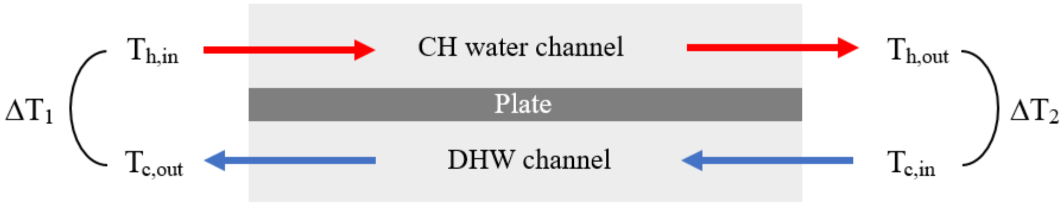


Figure 2.2: Inlet and outlet temperature modeling of the counter-flow PHE

The total heat transfer rate that is denoted as $\dot{Q}_{total}$ is determined from an energy balance of the water flowing through PHE channels. $\dot{Q}_{total}$ is calculated as shown in Equation (2.6) by using Equation (2.5). The calculation method is referenced in the study by Zhang et al. [6].

$$\dot{Q}_i = \dot{m}_i c_{p,i} \Delta T_i, i \in \{h, c\} \tag{2.5}$$

$$\dot{Q}_{total} = \frac{1}{2}\left(\dot{Q}_h + \dot{Q}_c\right) \tag{2.6}$$

Here, $\dot{m}$ is mass flow rate, c_p is the specific heat at constant pressure and ΔT is the temperature difference between the inlet and outlet temperatures of CH and DHW channels. The heat transfer rate in the hot channel and cold channel may differ in real cases, therefore the heat transfer rates are taken average to calculate the total heat transfer rate.

After calculating the total heat transfer rate and the LMTD with known heat transfer area, the overall heat transfer coefficient can be calculated as in Equation (2.7).

$$U = \frac{\dot{Q}_{total}}{A \Delta T_{lm}} \tag{2.7}$$

The particles accumulated on the surface of plates create an effect similar to an insulating layer. The fouling layer represents an additional resistance to heat transfer and causes the heat transfer to decrease. This effect of fouling on heat transfer is represented by a fouling resistance, R_f. The fouling factor is calculated for Test 2 to Test 8, shown in Table 2.1. The fouling factor is zero for Test 1, i.e., reference test conditions.

Heat transfer is carried out from the hot fluid, CH water, to the plate by convection, through the plate by conduction, and from the plate to the cold fluid, DHW by convection. The heat transfer to the surroundings is neglected while the PHE is modeling. The radiation effects are included in convection heat transfer coefficients.

The thermal resistance method is used to find the fouling resistance coefficient. The thermal network of a volume unit of PHE is shown in Figure 2.3. The convection heat transfer resistances, R_{CH} and R_{DHW}, that are generated from hot fluid to plate and plate to cold fluid and the conduction heat transfer resistance, R_P, that is generated through the plate also can be seen in Figure 2.3. The fouling resistance can be found in Equation (2.8) by using the convection, conduction resistance coefficient and overall heat transfer coefficient. In Equation (2.8), h_{CH} and h_{DHW} represent the convection heat transfer coefficient. A denotes the cross-sectional area. L is the length of the plate through the heat transfer direction. The conduction heat transfer coefficient denoted as k which is taken for the plate material, 316L stainless steel.

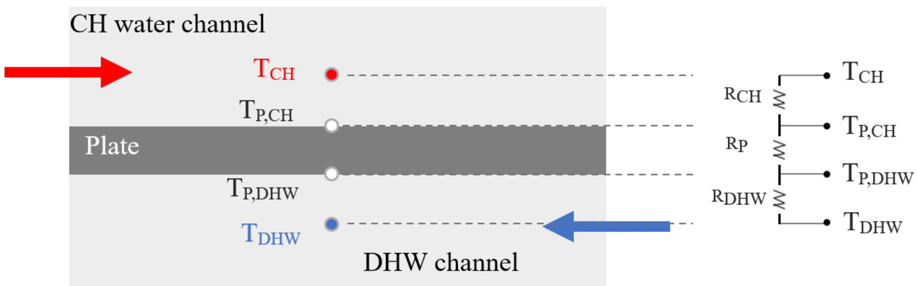


Figure 2.3: The thermal resistance network of a volume unit of PHE

$$\frac{1}{U.A} = R_{CH} + R_P + R_{DHW} + R_f = \frac{1}{h_{CH}A} + \frac{L}{kA} + \frac{1}{h_{DHW}A} + R_f \qquad (2.8)$$

2.2 Data Implementation of Fouling Assessment

The fouling effects on PHEs are investigated by generating a 1-D model in MATLAB program to validate the experimental results. The 1-D modeling is created based on finite volume elements. The PHE is modeled as discretized volumes. The PHE is modeled as two channels, CH and DHW, and one plate. The energy balance equation is used to model the heat transfer between the channels and the plate. The generated PHE model is discretized to volumes where the energy balance equation is applied. There is convective heat transfer between the channel and the plate for both hot and cold channels, i.e., CH and DHW channels, while there is conduction heat transfer from the hot surface of the plate where is faced to hot fluid to cold surface of the plate where is faced to cold fluid. This heat transfer flow is modeled based on thermal resistance method. The resistance model is created starting from the convection heat transfer from hot fluid to plate surface, then it goes with the conduction heat transfer through the plate, it is followed by the convection heat transfer from plate surface to cold fluid. Similar thermal resistance model is used with the one shown in Figure 2.3.

The fouling as mentioned in 2.1.4, is the accumulation of undesired particulates on plate surfaces. This accumulation results in generation of a fouling layer which acts like an insulation layer. Therefore, while the thermal resistance model is creating, the resistance of fouling layer, i.e., fouling resistance should be considered on the plate surfaces between the convection heat transfer coefficients from fluids to plate and conduction heat transfer coefficients through plate. The fouling resistance coefficients are added to the thermal resistance model as shown in Figures 2.4 and 2.5.

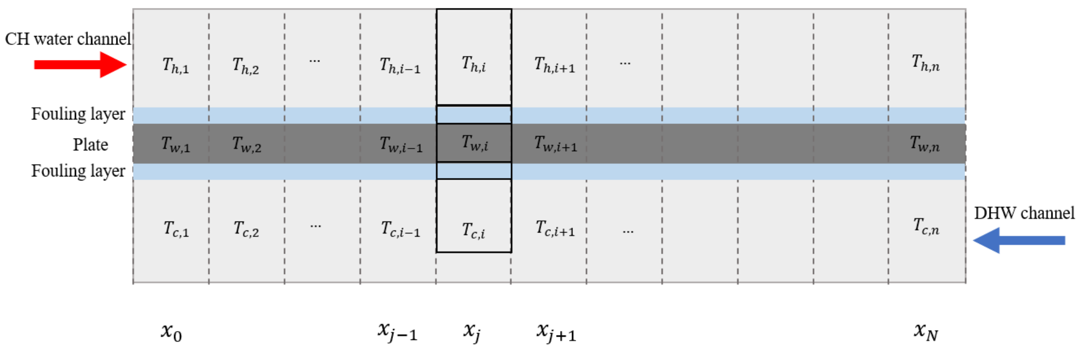


Figure 2.4. Discretized PHE model

The energy balance equations are created from a reference where the dynamic behavior of temperature distribution is studied by Bobic et al., [33]. In the equations of hot and cold channel (Equation (2.9) and (2.10)), left-hand side of the equations represents the internal energy that specified volume has. In the right-hand side of the equations, the heat transfer coming from the previous volume, the heat that is transferred to next volume and the heat transfer between the wall and fluid is modeled. The conduction heat transfer through the wall is modeled based on energy balance shown in Equation (2.11).

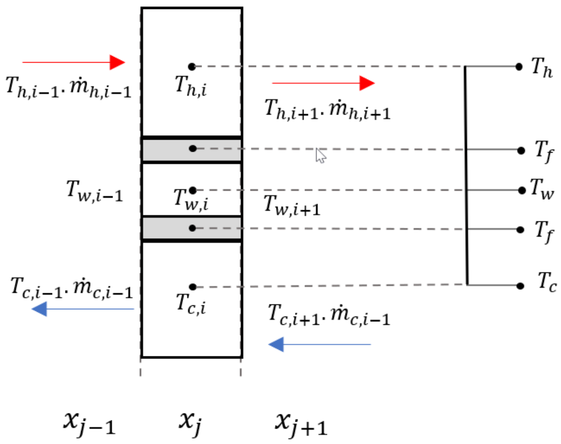


Figure 2.5. Thermal resistance model of PHE with the addition of fouling layers

$$V_h \rho c_p \frac{dT_{h,i}}{dt} = \dot{m}_h c_p T_{h,i-1} - \dot{m}_h c_p T_{h,i+1} - A\alpha_h \left(T_{h,i} - T_{w:h,i}\right) \qquad (2.9)$$

$$V_c \rho c_p \frac{dT_{c,i}}{dt} = \dot{m}_c c_p T_{c,i-1} - \dot{m}_c c_p T_{c,i+1} + A\alpha_c (T_{w:c,i} - T_{c,i}) \qquad (2.10)$$

$$V_w \rho c_p \frac{dT_{w,i}}{dt} = A\frac{2k}{l}\left[\left(T_{w:h,i} - T_{w,i}\right) - \left(T_{w,i} - T_{w:c,i}\right)\right] \qquad (2.11)$$

In the Equation (2.9), (2.10) and (2.11), V stands for volume, m^3, ρ represents the density of fluids, kg/m^3, T represents the temperature, °C, $\dot{m}$ represents the mass flow rate, kg/s, c_p represents the specific heat at constant pressure, J/kg°C, A represents the heat transfer area, m^2, α represents the convection heat transfer coefficient, W/m^2K, k represents the thermal conductivity, W/mK, and l represents the length of plate while the *h*,*c* and *w* are subscripts that stand for hot channel, cold channel and wall (plate), respectively.

The relation between surface of the wall and center of the wall is evaluated regarding the thermal resistance method as shown in Equation (2.12) and (2.13) to reduce the unknown variables. Here, U represents the overall heat transfer coefficient.

$$T_{w:h,i} = T_{h,i} - \frac{U_{h,i}}{\alpha_{h,i}}(T_{h,i} - T_{w,i}) \tag{2.12}$$

$$T_{w:c,i} = T_{c,i} - \frac{U_{c,i}}{\alpha_{c,i}}(T_{c,i} - T_{w,i}) \tag{2.13}$$

A relation between cell temperature, denoted as *i*, and previous and next cell temperature, denoted as *i-1* and *i+1*, is structured by taking average of the previous and next cell temperatures to indicate the cell temperature as shown in Equation (2.14) and (2.15).

$$T_{h,i} = \frac{T_{h,i-1} + T_{h,i+1}}{2} \tag{2.14}$$

$$T_{c,i} = \frac{T_{c,i-1} + T_{c,i+1}}{2} \tag{2.15}$$

By applying Equation (2.12), (2.13), (2.14) and (2.15) into Equation (2.9), (2.10) and (2.11), final equations used to structure the model in MATLAB given in Equation (2.16), (2.17) and (2.18).

$$\frac{dT_{h,i}}{dt} = C_1(T_{h,i-1} - T_{h,i}) - C_2(T_{h,i} - T_{w,i}) \tag{2.16}$$

$$\frac{dT_{c,i}}{dt} = C_3(T_{c,i-1} - T_{c,i}) + C_4(T_{w,i} - T_{c,i}) \tag{2.17}$$

$$\frac{dT_{w,i}}{dt} = C_5(T_{h,i} - T_{w,i}) - C_6(T_{w,i} - T_{c,i}) \tag{2.18}$$

$$C_1 = \frac{2\dot{m}_h}{V_h\rho}, C_2 = \frac{AU_{h,i}}{V_h\rho c_p} \tag{2.19}$$

$$C_3 = \frac{2\dot{m}_c}{V_c\rho}, C_4 = \frac{AU_{c,i}}{V_c\rho c_p} \tag{2.20}$$

$$C_5 = \frac{(A\frac{2k}{l} - \frac{A2kU_{h,i}}{l\alpha_{h,i}})}{V_w\rho c_p}, C_6 = \frac{(A\frac{2k}{l} - \frac{A2kU_{c,i}}{l\alpha_{c,i}})}{V_w\rho c_p} \tag{2.21}$$

In Equation (2.16), (2.17) and (2.18), the coefficients are indicated as C, where physical properties and constants are denoted. The coefficients are given in Equation (2.19), (2.20) and (2.21). The overall heat transfer coefficient, U, represents the combination of conduction, convection and fouling thermal resistance coefficients as shown in Equation (2.22) and (2.23).

$$U_{h,i} = (\frac{1}{\alpha_{\mathrm{h,i}}} + l/2k+R_{f,h})^{-1} \tag{2.22}$$

$$U_{c,i} = (\frac{1}{\alpha_{c,\mathrm{i}}} + l/2k + R_{f,c})^{-1} \tag{2.23}$$

The differential equations shown in Equation (2.16), (2.17) and (2.18) are solved by using Runge Kutta 4th order method. The Runge Kutta method is used for solving the ordinary differential equations (ODE). In solving methodology, the coefficients are calculated and the next step in time is calculated in the main equation (Equation (2.24)). The coefficient calculations are given in Equation (2.25), (2.26), (2.27) and (2.28).

$$y_{i+1} = y_i + 1/6(k_1 + 2k_2 + 2k_3 + k_4)h \tag{2.24}$$

$$k_1 = f(t_i, y_i) \tag{2.25}$$

$$k_2 = f(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h) \tag{2.26}$$

$$k_3 = f(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_2 h) \tag{2.27}$$

$$k_4 = f(t_i + h, y_i + k_3 h) \tag{2.28}$$

Here, in the equations, *h* stands for time step, *t* is time. The *y* represents the temperature of hot and cold channel, and plate in our calculations. The differential equation solvers are used for hot and cold channel, and plate.

2.3 Data Implementation of Partial Blockage Assessment

The assessment of partial blockage process is represented in titles of data grouping, data importing and model training.

2.3.1 Data Grouping

The supervised machine learning algorithms need a known set of input data grouped as predictors, i.e., parameters, and known responses to the data, like classes or labels. Here, during preprocessing the experiment data, 6 cases are created regarding to grouping the data by responses and predictors. The cases are listed in Table 2.2.

In 1st case shown in Table 2.2, the inlet and outlet temperatures, the pressure drops and the flow rates of CH and DHW channels are used as predictors, while the test conditions, i.e., Test 1 to 17 as shown in Table 2.1, are used as responses. Test 1 and Test 10 are representing the healthy value, therefore they grouped as Zone 0. Test 2 and Test 11 are representing the clogging of 2 plates, i.e., faulty value, thus they grouped as Zone 2. Similarly, Test 3 and 12 are Zone 3, Test 4 and 13 are Zone 4, etc. In the end, Test 9 is the only one in which the clogging of 16 plates effects is tested and represented as Zone 8.

In 2nd, 3rd and 4th cases, the calculated overall heat transfer coefficient (U) is used as predictor. In 3rd case, to reduce the group number, the test results corresponding to the test conditions, i.e., Zone 0 to 5, are grouped in 2, e.g., results of Zone 0 and Zone 1 are grouped as Group 1, etc. The final group is created for Zone 6,7 and 8. The clogging plate numbers in the tests, can be represented as clogging percentage as well. The maximum clogging level indicates 50% clogging in PHEs for evaluating the clogging of 16 plates of 32 plates PHE. The clogging percentages are calculated for each test and are represented in 4th cases as percentage groups, from 0-10% clogging to 40-50% clogging.

Table 2.2: Experiment data grouping cases according to predictor and response group

Case	Predictor	Response Group
1	• CH inlet temperature • DHW inlet temperature • CH outlet temperature • DHW outlet temperature • CH flow rate • DHW flow rate • CH pressure drop • DHW pressure drop	• Zone 0 • Zone 1 • Zone 2 • Zone 3 • Zone 4 • Zone 5 • Zone 6 • Zone 7 • Zone 8
2	• Overall heat transfer coefficient	• Zone 0 • Zone 1 • Zone 2 • Zone 3 • Zone 4 • Zone 5 • Zone 6 • Zone 7 • Zone 8
3	• Overall heat transfer coefficient	• Group 1 • Group 2 • Group 3 • Group 4
4	• Overall heat transfer coefficient	• 0-10%, P1 • 10-20%, P2 • 20-30%, P3 • 30-40%, P4 • 40-50%, P5
5	• CH pressure drop • DHW pressure drop	• Zone 0 • Zone 1 • Zone 2 • Zone 3 • Zone 4 • Zone 5 • Zone 6 • Zone 7 • Zone 8

Table 2.2 (continued): Experiment data grouping cases according to predictor and response group

6	• CH inlet temperature • DHW outlet temperature • DHW flow rate	• Zone 0 • Zone 1 • Zone 2 • Zone 3 • Zone 4 • Zone 5 • Zone 6 • Zone 7 • Zone 8

In 5^{th} case, the pressure drops of CH and DHW channels are used as predictors to see if the only parameters known would be the pressure drops which would be the performance of the machine learning algorithms classification. In similar, the CH inlet temperature, DHW outlet temperature and DHW flow rate are used as predictors in the 6^{th} case to see if the algorithm performance would be sufficient with these parameters. These parameters are selected because they are measured during real life operation in combi-boiler.

The test conditions are representing the fouling levels as they are representing the customer comfort. Therefore, the test conditions and the categorized levels, e.g., Zone 0 to 8, show the customer comfort levels. The higher the heat transfer rate, the higher the customer would achieve the desired comfort. Thus, the comfort is decreasing while the fouling is increasing. The zone categorization shown in Figure 2.6, is representing the comfort levels corresponding the zones. Zone 0, as Test 1 and 10, indicates the zero-hour performance of the PHEs, i.e., there is no fouling on plate surfaces. The customer comfort and heat transfer rate are at maximum in Zone 0; thus, the categorized region is named full comfort. While the fouling is increasing, the comfort loss starts. In Zone 1 and 2, the status of fouling is represented as mostly comfort region. Here, there are unnoticeable effects on the heat transfer rate by customer yet. The 3^{rd} region is named partial comfort loss which is associated with Zone 3 and 4. Here, customer may notice a lack of heat transfer rate. There may not have the same performance to heat the DHW when compared to the zero-hour performance anymore. In the 4^{th} region, there are serious comfort loss corresponding to Zone 5 and 6, that customers can notice. Here, the performance of PHE is likely to be affected seriously by fouling on plate surfaces. In the last region, partial clogging is expected in the plates of PHE that can affect the heat transfer rate extremely. In this point, the PHE might need to change not to affect and prevent the required customer comfort and it is represented with the categorized Zone 7 and 8.

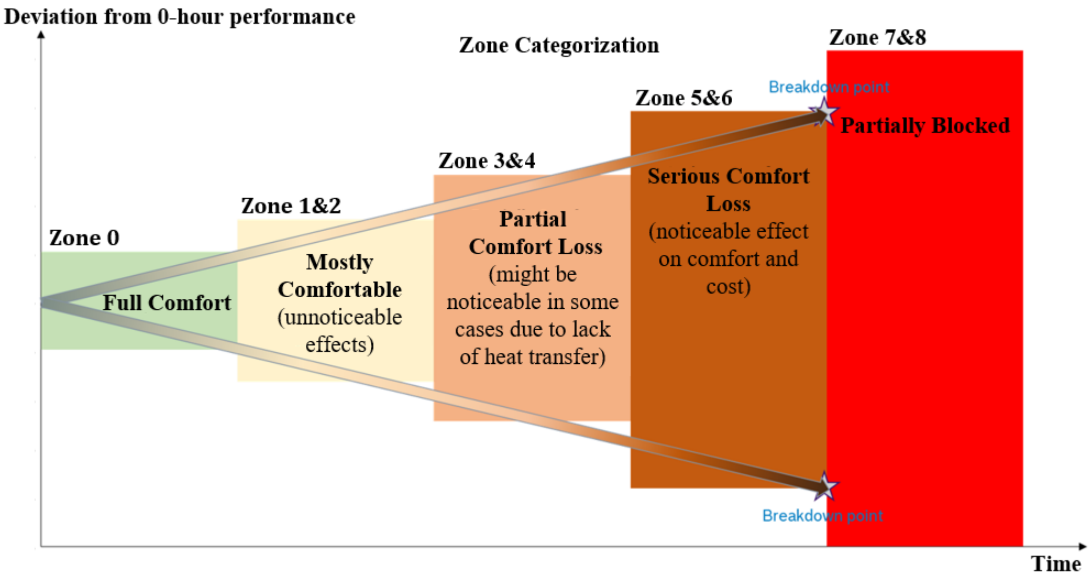


Figure 2.6: The zone categorization of customer comfort regarding fouling

2.3.2 Data Importing

During classification, the algorithm should find the target class for a new data sample that is not categorized yet, given a set of training data and corresponding training classes [25]. The classification process has two steps. One is training and the second one is the testing step. In the training step, a model is constructed from the training data that is generated from the experiments. In the testing step, the constructed model is used to classify the test data which is taken from experiments also.

The experiment data is imported into the programming platform called MATLAB. The classification process is conducted in the Classification Learner App inserted in MATLAB. Classification Learner App is used for training the models with given training data to classify new data, i.e., supervised machine learning. Here in the app, the validation schemes, models to be trained and result assessment tools have existed.

The data splitting for training and testing data has crucial points. During this splitting process, the k-fold cross validation technique is used. The k-fold cross validation technique is commonly used for avoiding overfitting or underfitting. The k-fold cross validation splits the data into k number subsets, i.e., folds, in equal size. In each fold, one part of data is used to train the model, and the other part of the data is used to test the model. And in each fold, the testing data part is another 20% parts of the main data. The partition of data by k-fold cross validation techniques is shown in Figure 2.7. Each fold runs and obtains a learning accuracy for each fold. The final prediction accuracy of the used model is calculated by averaging the learning accuracies obtained for each fold. The fold number, k, is chosen as 5 in this study.

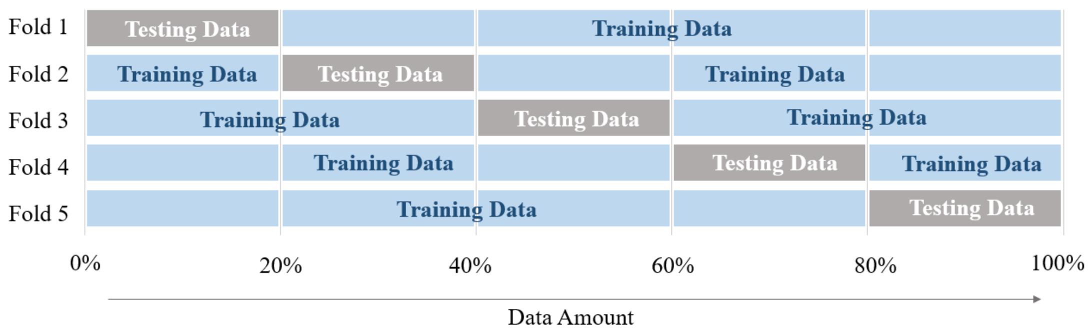


Figure 2.7: The k-fold cross validation representation

2.3.3 Model Training

Each classification model is unique with its strength and weaknesses regarding the case where it would be used. Choosing the right model generally requires a trial and error method to get the balance of performance and accuracy. Therefore, 3 models, Naïve Bayes, decision tree, and k-nearest neighbors are chosen in this study to investigate their performance and accuracy.

The Naive Bayes classifier is a straightforward probabilistic classifier that uses the Bayes theorem along with strong (naive) independence assumptions. The Bayes theorem is used to obtain the posterior probability within the Naïve Bayes classifier. The posterior probability is the probability of which class a particular data may belong to, while the probability of selecting a particular data from a class is called prior probability. In the training step of Naïve Bayes, with the given class and data, first, the prior probabilities are calculated, then the posterior probability will be the output for possible classes. If we assume the classes, in this study, are named zones, are denoted as *y* and the features (predictors) are denoted as *x*. The main task in Naïve Bayes classification is giving the posterior probability, i.e., the probability of which *y*, a *x* may belong, using the Bayes theorem shown in Equation (2.28) [25]. Here the *k* denotes the random variable corresponding to classes (*y*), *j* denotes the random variable corresponding to predictors (*x*), *p* denotes the number of predictors (*x*), and *K* denotes the number of classes (*y*). $P(y = k|x_1, \cdots x_P)$ denotes posterior probability, while $P(x_j|y = k)$ denotes prior probability. $P(y = k)$ denotes the marginal probability which is the probability of selecting the class among the total classes.

$$P(y = k|x_1, \cdots x_P) = \frac{P(y = k)\prod_{j=1}^{p} P(x_j|y = k)}{\sum_{k=1}^{K} P(y = k)\prod_{j=1}^{p} P(x_j|y = k)} \qquad (2.28)$$

Naïve Bayes classifier is a model that has high bias and low variance characteristics. High bias refers to the error between the real class and variance refers to the ability to achieve approximate accuracy with different training sets. This characteristic of Naïve Bayes classifier provides decreasing in risk of inaccurate predictions but has the probability of not properly matching the data set to the model. The Naïve Bayes model is supported by various distributions in Classification Learner App. The kernel distribution is selected. Therefore, after that the chosen model would be named Kernel Naïve Bayes in this study. The kernel distribution is a function that is used in non-parametric estimations. Non-parametric estimators do not have a defined structure and rely on all data points to conclude. To see if the performance of the algorithm that would be placed in a combi-boiler during operation in real life would be sufficient or not with independently read parameters, the kernel distribution is used. There are several kernel functions that MATLAB provides as inaccessible content. In this study, triangle named kernel distribution is used.

Decision trees are used for both classification and regression. Here the usage of the decision tree algorithms is determined as classification. This model also can be called as classification tree. A decision tree can be considered as a predictor that predicts the class corresponding to an instance data by partitioning the given training data into the labeled data. It is shown as a road from root to leaf and nodes in between. An example of a decision tree can be seen in Figure 2.8. At each node on the paths, a selection of a response according to the condition is generated until it reaches the final classified leaf. The classification decision trees are binary. So, each step in a prediction, checking the condition and deciding a response as 1 for true or 0 for false are involved in the process [26].

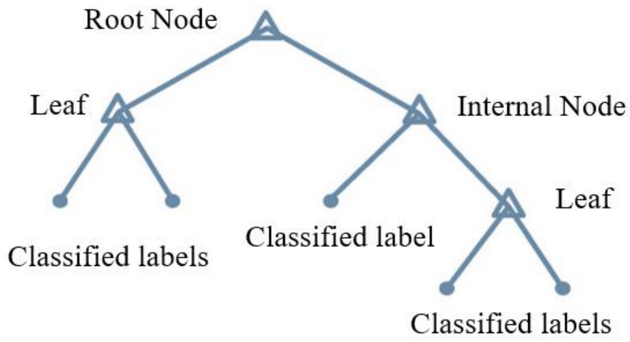


Figure 2.8: Decision tree example and tree designation

During the classification process, many algorithms is used to determine the creation of the tree. CART (Classification and Regression Trees) is used in this study as mentioned due to classification output being required. The second step that has crucial impact on the creation of the tree, is determining the attributes. There are various attribute selection techniques, i.e., splitting criterion, the Gini index which is used in this study is quite popular in the literature [27]. A decision tree determines how to split nodes either according to impurity or node error. The Gini index is based on impurity. A node with only one class, i.e., a pure node, has a Gini index of 0, while other nodes are having a positive Gini index. Thus, the Gini index would be a measure of node impurity. The impurity The Gini index splitting criterion can be expressed as the error that would be encountered if each item were categorized at random using the probability distribution of class labels within each subgroup [25]. The Gini index can be found as shown in Equation (2.29). Here, for any node x and class y, with total class number k, $p_y(x)$ denotes the probability of an instance being classified to a particular class [28].

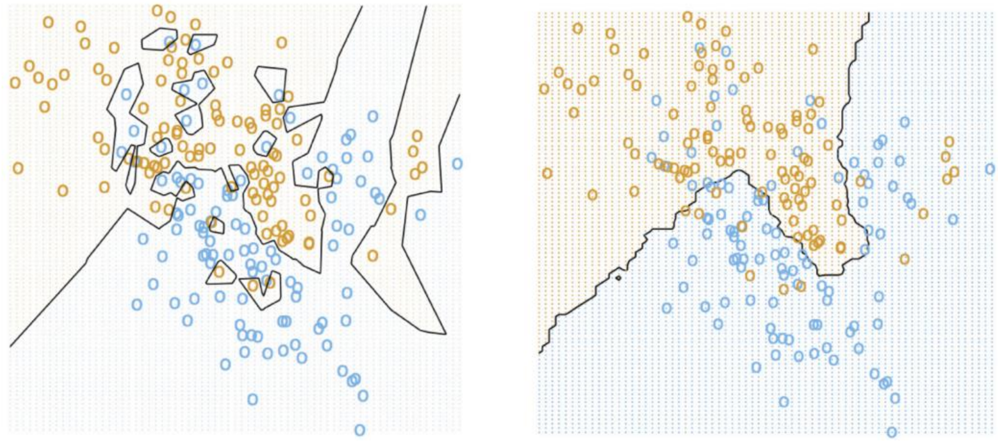
$$Gini(x) = 1 - \sum_{y=1}^{k} \left(p_y(x)\right)^2 \tag{2.29}$$

The k-nearest neighbors (kNN) classifier model provides a label based on the dominant class in the region, i.e., neighborhood, by locating a cluster of k training set objects that are closest to the test data. The two fundamental elements of this approach are a distance metric to estimate distance between objects and the number of nearest neighbors which is denoted as k to classify unlabeled data. To determine its nearest-neighbor list, I_z, the kNN model calculates the similarity distance between a training set, $(x, y) \in I$, and test data, $z = (\hat{x}, \hat{y})$. Here, *x* represents the training instance, and *y* represents the corresponding class, while $\hat{x}$ and $\hat{y}$ represent the test instance and its class, respectively. The algorithm can be shown in the steps listed below [27]:

- Import training set (x, y), and testing set $z = (\hat{x}, \hat{y})$ as inputs.
- Calculate the distance $d = (\hat{x}, x)$, between each instance in training data and testing data.
- Select the nearest neighbor list, $I_z \subseteq I$, which is the set of k closest training instances to testing instances, *z*.
- Compute the output, i.e., the testing classes which are the target, $\hat{y}$, by using Equation (2.30).

$$\hat{y} = arg_v max \sum_{(x_i, y_i) \in I_z} F(v = y_i) \tag{2.30}$$

Here in Equation (2.11), $F(v)$ will be resulted as 1 if the argument, *v*, is true and 0 otherwise. The *v* is the class label [27]. The operation of the kNN model highly depends on the *k* number selection, which is mentioned in the third step of the process. Choosing the k number as 1, results in 0 error and 100% accuracy, due to it being classified as itself [29]. This is not a required solution. By choosing this, the model will be overfitted to the training and the trialed test set and provides a very low, nearly zero bias. This results in an increase in the dependence of the model on the selected test and training data set, i.e., the variance of the model will be too high. The optimal result of the model that is tried to be achieved is high bias but low variance characteristic. To obtain this required solution by kNN model, the k number should be increased [29].



(a) k=1 (b) k=15

Figure 2.9: Voronoi diagram of kNN algorithm for k=1 and k=15 [30]

As can be seen in Figure 2.9, the kNN model sets the nearest neighbor list according to k number. The shown orange and blue classes are classified by kNN algorithm for k=1 and k=15 in Figure 2.9. As mentioned, when the k number is selected as 1, the classes are distinguished as too specific to the training data, whereas when the k number is selected as higher than 1, the bias is increasing yet the variance, i.e., sensitivity to the training set is decreasing. But at one point, if k is selected as too large, the model will underfit the training data, i.e., the bias will be too high that cannot fit almost the training data at all [31]. In this study the k number is chosen as 10.

In the mentioned process step, which is computing the k closest training data points to testing data, the closeness can be quantified with distance functions. The Euclidean distance is used. As it is mentioned in the second step of the process, the distance d is calculated by the Euclidean distance method, which is just computing the tangent distance between each test data instance and each training data instance. The Euclidean distance is calculated by using the basic expression shown in Equation (2.31) [30].

$$d_i = \|x_i - \hat{x}_i\| \tag{2.31}$$

Chapter 3

Results and Discussion

3.1 Experiment Results

The 30 and 32 plates PHEs are tested by measuring pressure drops and outlet temperatures of both channels, CH and DHW whereas inlet temperatures and volume flow rates are given as input. The results are grouped as shown in the cases listed in Table 2.2. As given in Table 2.2., there are 3 different groups for responses, Zone 0 to Zone 8 grouping, grouping zones by 2 and clogging percentages grouping. These response cases are rearranged by normalizing the data. For Zone 0 to 8 grouping, the zones are considered as time as they are the representation of fouling and clogging by time. So, the zones are normalized between 0 to 1. The procedure is applied to the other response grouping cases as well.

During experiments, 300-500 pieces of data are acquired for each test. The results data that is obtained from each test condition are grouped as same and represented in Figure 3.1 and 3.2 as a data group colored in the same. In Figure 3.1 the experiment results of 30 and 32 plates are represented. As expected, while the clogging plate number that is simulated with tests is increasing (shown in normalized time), the CH outlet temperature is increasing while the DHW output temperature is decreasing. The increase of CH outlet temperature is a demonstration of decreasing heat transfer. The heat that the CH inlet flow has cannot be transferred as in the zero-hour performance. Similarly, the DHW output has not gained the heat as in the zero-hour performance which is the reason gives the lack of comfort to customers. While fouling and clogging are increasing, the layer of fouling creates hedges to the flow. This results in an increasing pressure drop. The pressure difference results show the expected increase while clogging levels are increasing.

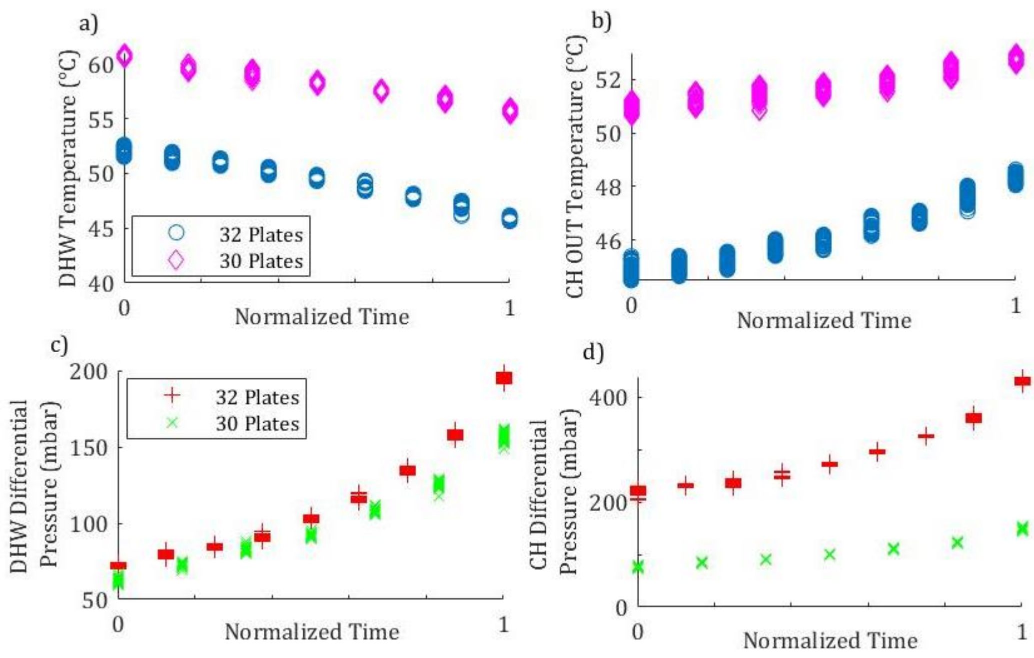


Figure 3.1: The experiment results of 32 and 30 plates PHEs in normalized time; a) DHW outlet temperature, b) CH outlet temperature, c) pressure drop of DHW channel, d) pressure drop of CH channel.

The overall heat transfer coefficient for each data is calculated as in Equation (2.7). The results are shown in Figure 3.2 and 3.3 for 32 and 30 plates PHEs, respectively. The decreasing trend is a representation of the decreasing heat transfer as expected. The overall heat transfer coefficients of 32 plates are larger than the coefficients of 30 plates as shown in the graphs.

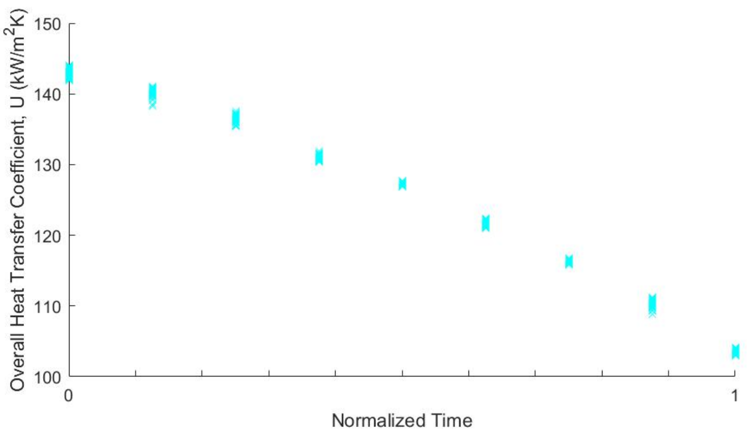


Figure 3.2: The calculated overall heat transfer coefficient for 32 plates PHE in the normalized time scale

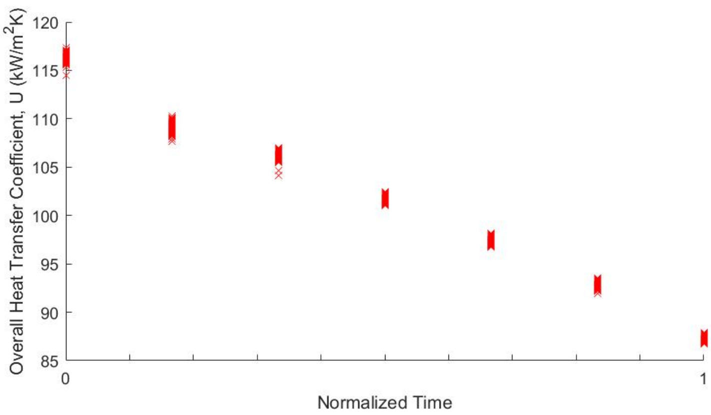


Figure 3.3: The calculated overall heat transfer coefficient for 30 plates PHE in the normalized time scale

During the fouling resistance coefficient evaluation, the required convection heat transfer coefficient calculation is carried out by using the Sieder-Tate Nusselt number correlation, as shown in Equation (3.1) [32]. This Nusselt correlation can be used when the conditions are given in Equation (3.2), (3.3) and (3.4) [32]. The Prandtl number, Reynolds number and dynamic viscosities are given in Table 3.1. The value of the length of the plate over the hydraulic diameter is 54.5 for 32 and 30 plates PHEs.

$$Nu = 0.27Re^{0.8}Pr^{1/3}(\mu_f/\mu_w)^{0.14} \tag{3.1}$$

In Equation (3.1), The *Nu* represents the Nusselt number, *Re* is the Reynolds number, Pr is the Prandtl number, μ_f is the dynamic viscosity of the fluid at fluid film temperature, i.e., water and μ_w is the dynamic viscosity of the fluid at wall temperature, i.e., plate. The wall and initial temperatures are accepted as the same as the cold channel water inlet temperature when the model is structured. The film temperature is calculated by taking the average of the cold channel and hot channel inlet temperatures.

$$0.7 \leq Pr \leq 16 \tag{3.2}$$

$$Re \geq 10{,}000 \tag{3.3}$$

$$L/D \geq 10 \tag{3.4}$$

Table 3.1: The properties of the hot and cold channel

		Hydraulic diameter (m)	Reynolds Number	Prandtl Number at the film temperature	Dynamic viscosity (Pas) at the film temperature	Dynamic viscosity (Pas) at plate temperature
32 Plates PHE	Hot channel	0.0035	2.7×10^5	4.495	0.652×10^{-3}	1.306×10^{-3}
	Cold channel	0.0035	1.6×10^5	4.495	0.652×10^{-3}	1.306×10^{-3}
30 Plates PHE	Hot channel	0.0035	2.3×10^5	4.495	0.652×10^{-3}	1.306×10^{-3}
	Cold channel	0.0035	0.9×10^5	4.495	0.652×10^{-3}	1.306×10^{-3}

The Reynolds number is calculated by Equation (3.5). The ρ denotes the density, V denotes the velocity of fluid, D_h denotes the hydraulic diameter, and μ denotes the dynamic viscosity of fluid. The density and dynamic viscosities are taken at film temperature. The velocities are derived from the volume flow rates which are calculated by using the geometric properties of PHEs and mass flow rates. The hydraulic diameter is calculated by Equation (3.6), where A_c is cross-sectional area and p is the wetted perimeter of channel. The hydraulic diameter and relatively Reynolds number are calculated for both hot and cold channel of 32 and 30 plates PHEs, as shown in Table 3.1.

$$Re = \frac{\rho V D_h}{\mu} \tag{3.5}$$

$$D_h = \frac{4A_c}{p} \tag{3.6}$$

After calculating the Reynolds number and with known Prandtl number and dynamic viscosities the Nusselt numbers are calculated by using Equation (3.1), the values of its given in Table 3.2.

Table 3.2: The Nusselt number calculated by correlation for hot and cold channel

	Nu calculated analytically by correlation	
	32 Plates	30 Plates
Hot channel	8.91×10^3	7.85×10^3
Cold channel	6.12×10^3	3.88×10^3

The convection heat transfer coefficients are calculated by Equation (3.7) where the k is thermal conductivity of fluid and D_h is hydraulic diameter. The thermal conductivity of fluid is taken at film temperature. The convection coefficients that are calculated analytically by Nusselt number correlation are listed in Table 3.3.

$$h = \frac{Nuk}{D_h} \tag{3.7}$$

Table 3.3: The convection heat transfer coefficients

	Convection heat transfer coefficient calculated analytically (W/m^{2}K)	
	32 Plates	30 Plates
Hot channel	1.01×10^6	1.41×10^6
Cold channel	1.60×10^6	0.69×10^6

After the convection heat transfer coefficients are calculated, the conduction coefficient of plate is also calculated as designated in Equation (2.8). The found result is 0.00022 W/mK.

The fouling resistance coefficients that are calculated are shown in Figure 3.4 and 3.5 for 32 and 30 plates PHEs, respectively. As expected, when the decreasing overall heat transfer coefficient is considered while the fouling is increasing, the fouling resistance coefficients are increasing. These data are also represented in the normalized time scale as it is in the representation of the overall heat transfer coefficient.

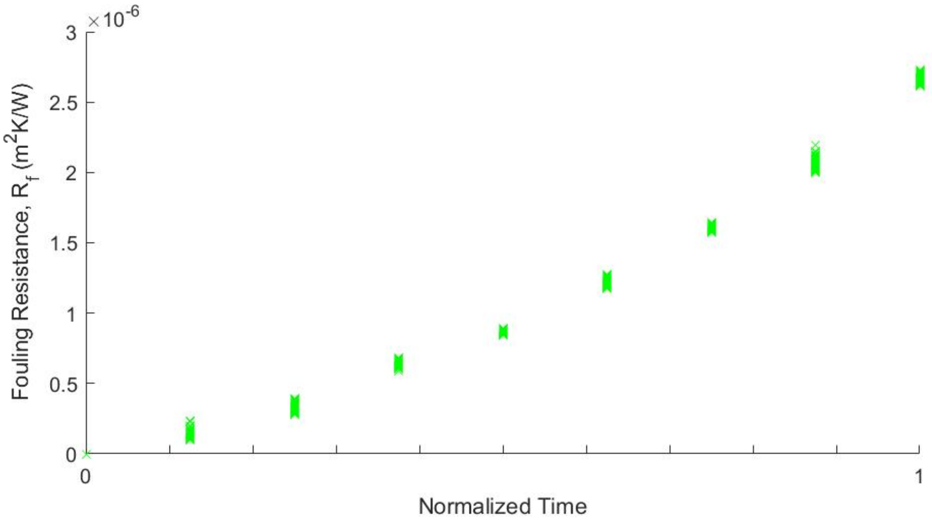


Figure 3.4: The calculated fouling resistance coefficients of 32 plates PHE

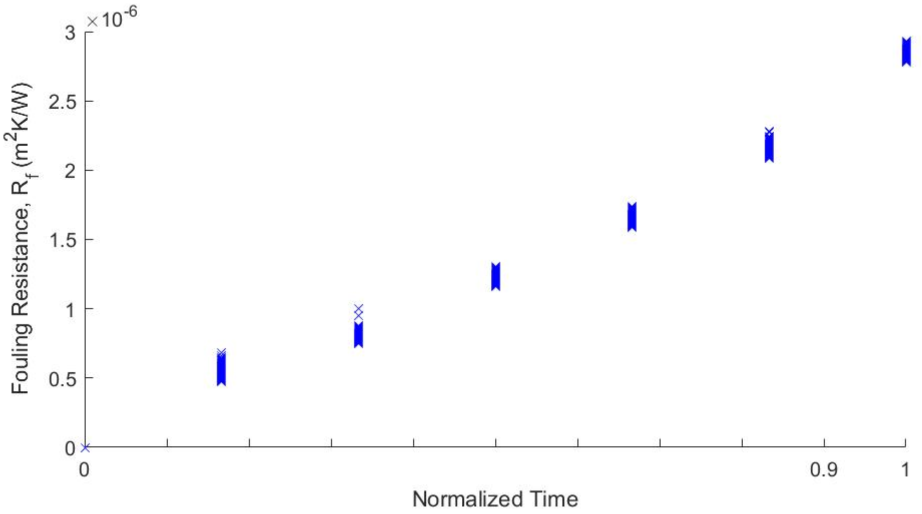


Figure 3.5: The calculated fouling resistance coefficients of 30 plates PHE

The reference study evaluated by Zhang et.al. shows the result of the fouling resistance values with time [6]. The used method to create the fouling environment in PHEs is adding the particulates into the test rig. Therefore, an accelerated fouling effect observation is carried out in the study. There are two types of fouling tested, one is particulate fouling which only the Al_2O_3 accumulation is observed and the other one is composite fouling which the $CaCl_2$ and $NaHCO_3$ precipitation is observed together with Al_2O_3. In both tests, the fouling resistance results show both the asymptotic increasing and parabolic increasing. The fouling resistance coefficients in this study show a similar increase to the results shown in the study of Zhang et.al [6]. However, the fouling resistance values show differences between the referenced study and this study. The average of the calculated fouling resistance values is approximately $2x10^{-6}$ m^2K/W, while the magnitudes of fouling resistance coefficients in the referenced study are the value times 10^{-4}. This difference originated from the test method differences between the studies. The accelerated fouling test shows a much larger accumulation and lack of heat transfer than the results found from the generated artificial fouling test method in this study.

3.2 Numerical Model Results

The model is structured in MATLAB, by using Runge Kutta 4^{th} order method as mentioned. The Runge Kutta 4^{th} order method is referenced from the study by Bobic et al. [33]. Therefore, once the model is created, the parameters are used to validate the model with a reference study. The temperature of the hot channel inlet is given in referenced study as 52.4°C and temperature of the cold channels is given as 18.5°C. Both hot and cold channel have the same mass flow rate, which is 0.113 kg/s. The initial temperature of the plate is assumed as the same temperature as the cold channel inlet temperature. The step time of model is stabilized by running the model for the parameters of the volume elements number.

In the reference study of the model results obtained [33], the published hot channel temperature distribution was seen to have similar behavior. For the initial and boundary conditions used in the reference study given above, it was reported that the system reached stable conditions within 5 seconds after starting from the initial status of the system and the hot channel output temperature was approximately 27 ° C in the experimental results performed with the thermal camera [33]. In the validation study, the hot channel outlet temperature for the same time was found to be 29 ° C (Figure 3.6). Time-dependent temperature distribution is compatible with the model results. As a result, reached steady state status is a parameter indicating the accuracy of the model (Figure 3.7). In the validation of the reference study, the time-related graph and location of the hot and cold channel temperatures obtained were given (Figure 3.8). The number of locations indicated in Figure 3.8 on the Y axis is determined by leaving the cell with the starting limit and the graph was formed. According to this validation study, it was concluded that the model runs correctly.

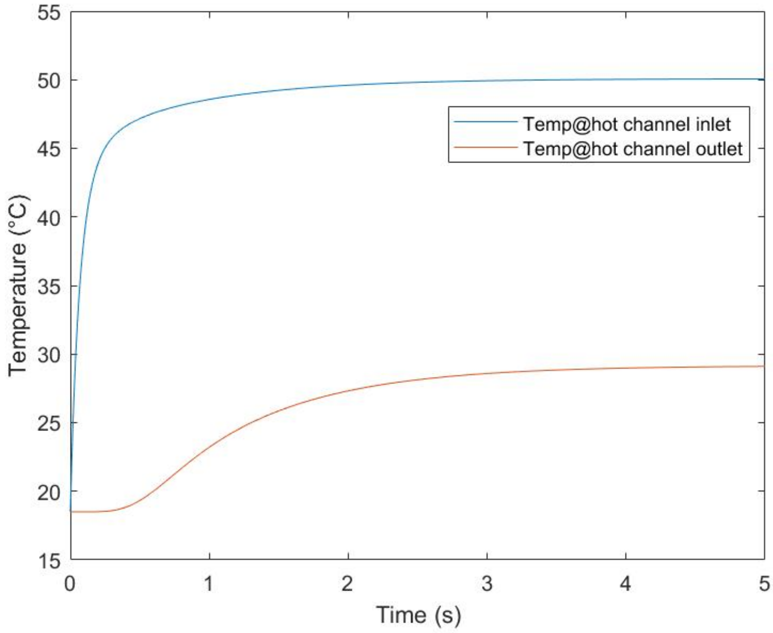


Figure 3.6: Temperature distribution of hot channel inlet and outlet by time in validation of reference study

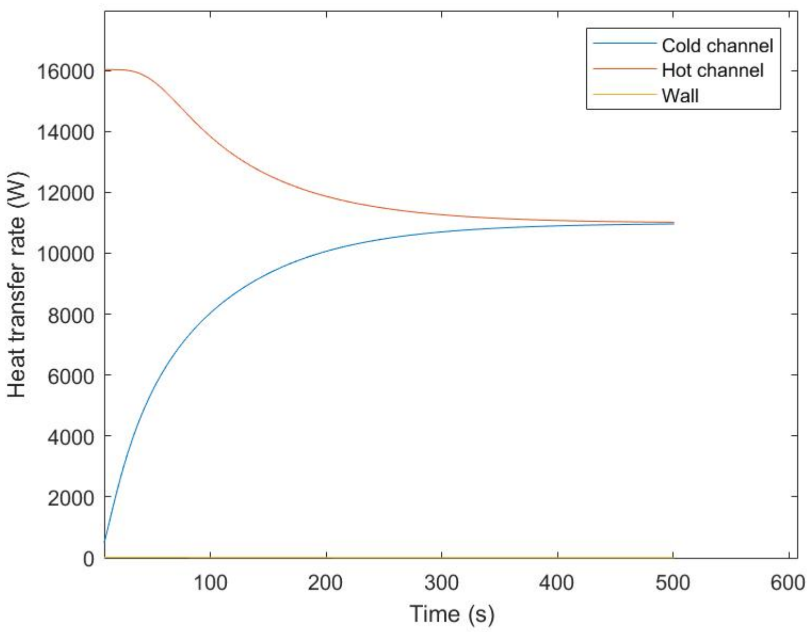
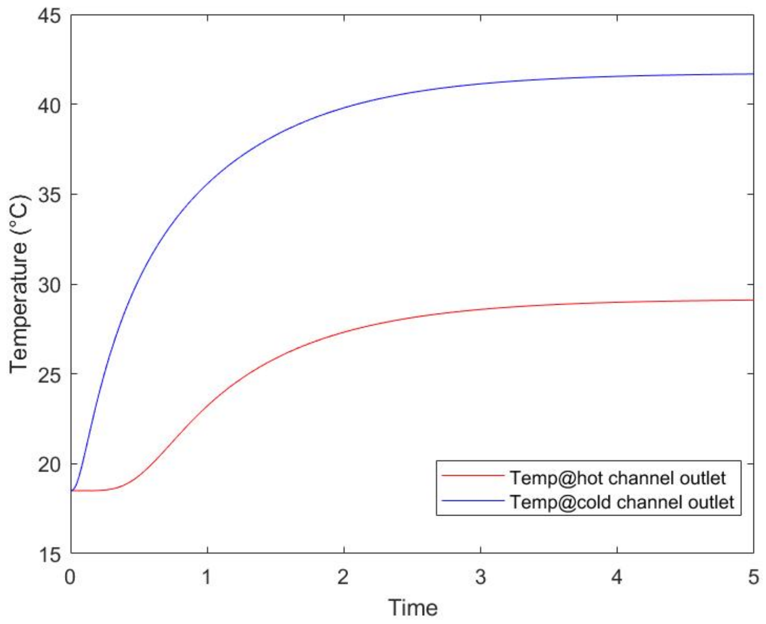
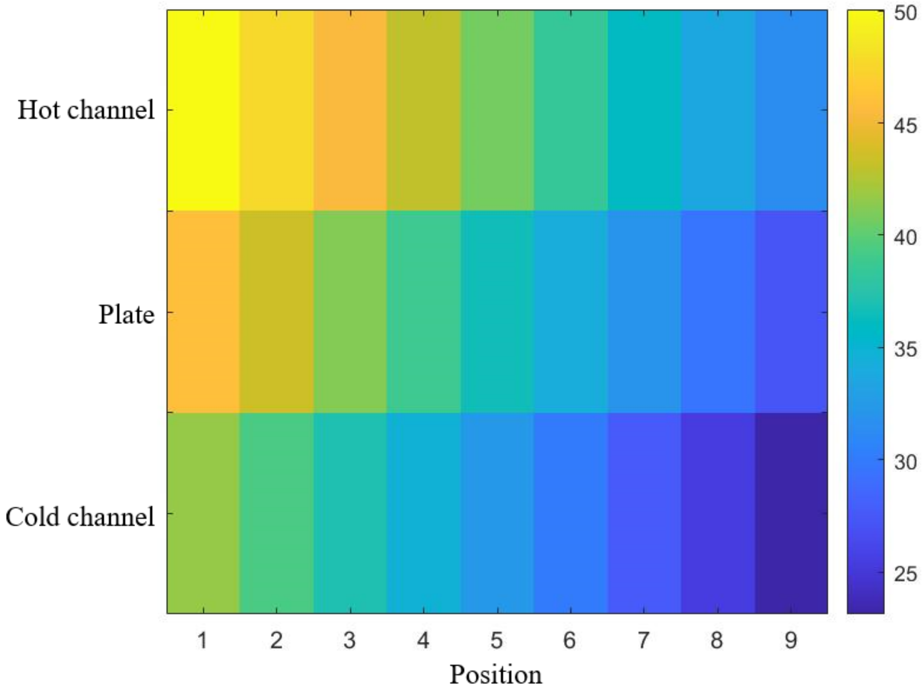


Figure 3.7: Heat transfer rate change by time of hot and cold channel in validation of reference study



(a)



(b)

Figure 3.8: Temperature distribution of a) hot and cold channel by time, b) hot and cold channel, and plate by position

After validation study, the parameters of PHEs have 32 and 30 plates applied to the model. In the first part of the study, the healthy conditions were examined by using the 32 and 30 plate PHEs flow rates and inlet temperatures. As in the validation study, steady state status is reached as shown in Figure 3.9 and 3.10 for 32 and 30 plate PHE, respectively.

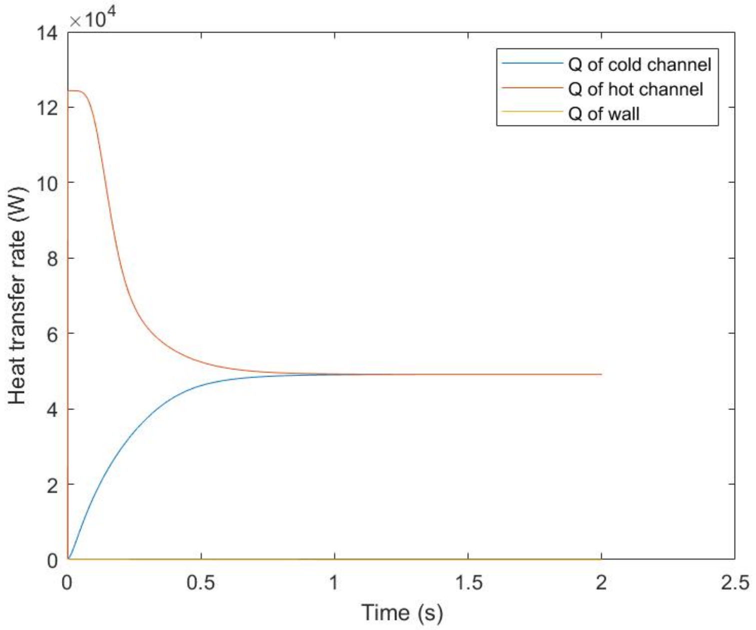


Figure 3.9: Heat transfer rate change by time for 32 plates PHE

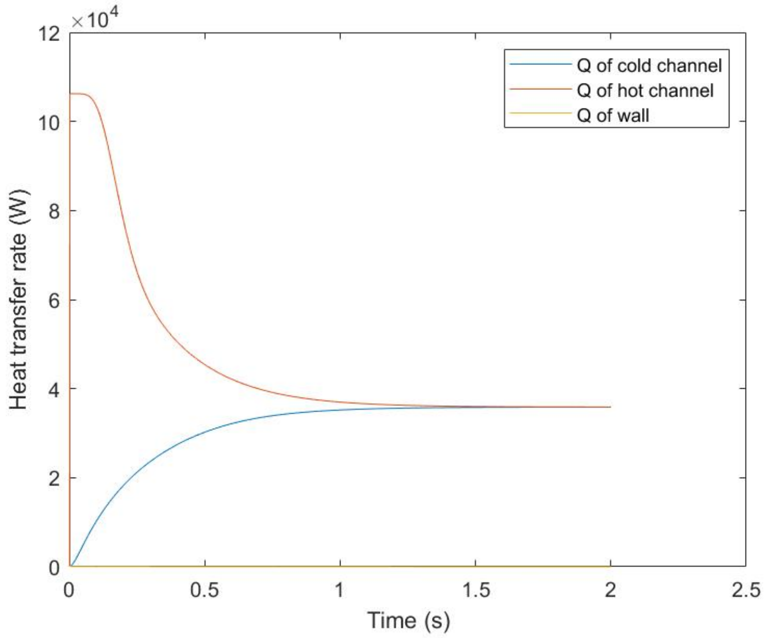


Figure 3.10: Heat transfer rate change by time for 30 plates PHE

Different Nusselt number correlations are used to create the optimum model that gives similar results when it is compared with the test results. The used Nusselt number correlations are given in Table 3.4. Nusselt number correlation 1 is derived from Nusselt number correlation 3 [32] to obtain the optimum model results. The correction factors are changed due to different Reynolds number values of hot and cold channels.

Table 3.4: The Nusselt number correlations and corresponding convection heat transfer coefficients used in the model

		Nusselt number correlations	Convection heat transfer coefficient (W/m^2K) 32 Plates PHE	Convection heat transfer coefficient (W/m^2K) 30 Plates PHE
1	Hot channel	$Nu_h = 0.15Re^{0.8}Pr^{\frac{1}{3}}(\mu_f/\mu_w)^{0.14}$	8.9×10^5	7.8×10^5
	Cold channel	$Nu_c = 0.08Re^{0.8}Pr^{\frac{1}{3}}(\mu_f/\mu_w)^{0.14}$	3.26×10^5	2.1×10^5
2	Hot channel	$Nu = 0.023Re^{0.8}Pr^{0.3}$	1.4×10^5	1.26×10^5
	Cold channel		0.98×10^4	6.23×10^4
3	Hot channel	$Nu = 0.27Re^{0.8}Pr^{\frac{1}{3}}(\mu_f/\mu_w)^{0.14}$	1.6×10^6	1.4×10^6
	Cold channel		1.1×10^6	6.9×10^5
4	Hot channel	-	59430	58703
	Cold channel			

The 2nd case of Nusselt number correlation is commonly used Dittus-Boelter Nusselt number correlation [32]. In the 4th case, the convection heat transfer coefficients that are taken from CFD simulations are used. The CFD simulations were carried out as a background study in Bosch Thermotechnology. The details of the study would not be appropriate to be shared due to the legislation of the patent taken by the company itself.

The convection heat transfer coefficients are calculated as an average area weighted in 3D CFD simulations of whole body of PHE. The results of the used correlations are represented in Figure 3.11 and Figure 3.12. The outlet temperature of the hot and inlet channel is depicted in the figures. The 3rd case, Dittus-Boelter Nusselt number correlation, shows similar outputs with test results. However, the derived Nusselt number correlation, 1st case, shows better results with the test results. The outermost result is seen in the 4th case, when the convection heat transfer coefficients taken from CFD results are used. The 1st case results are similar for 32 and 30 plates PHEs. Therefore, the derived Nusselt number correlations are selected to be used for the further model calculations.

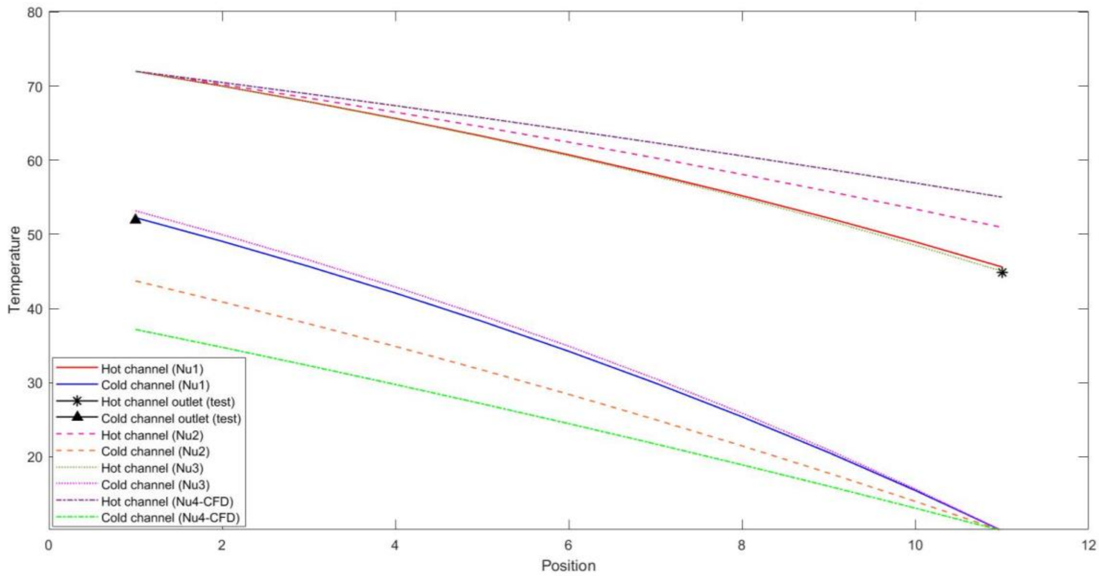


Figure 3.11: The Nusselt number correlation comparison for 32 plates PHE

After selection of the optimum Nusselt number correlation, the model results for healthy conditions are shown in Figure 3.13, 3.14, 3.15, 3.16, 3.17 and 3.18. The temperature distribution of wall (plate), hot and cold channels are shown. Again, the outlet temperatures of hot and cold channels from test results are indicated in the figures. The results show that the model can give compatible temperature distribution evaluation with test results for 32 and 30 plates PHEs.

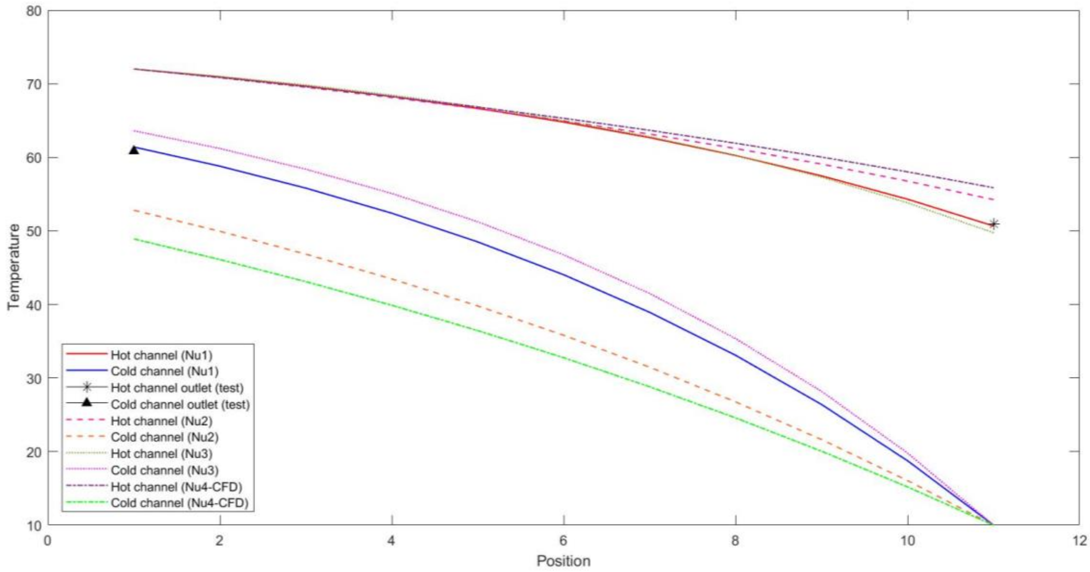


Figure 3.12: The Nusselt number correlation comparison for 30 plates PHE

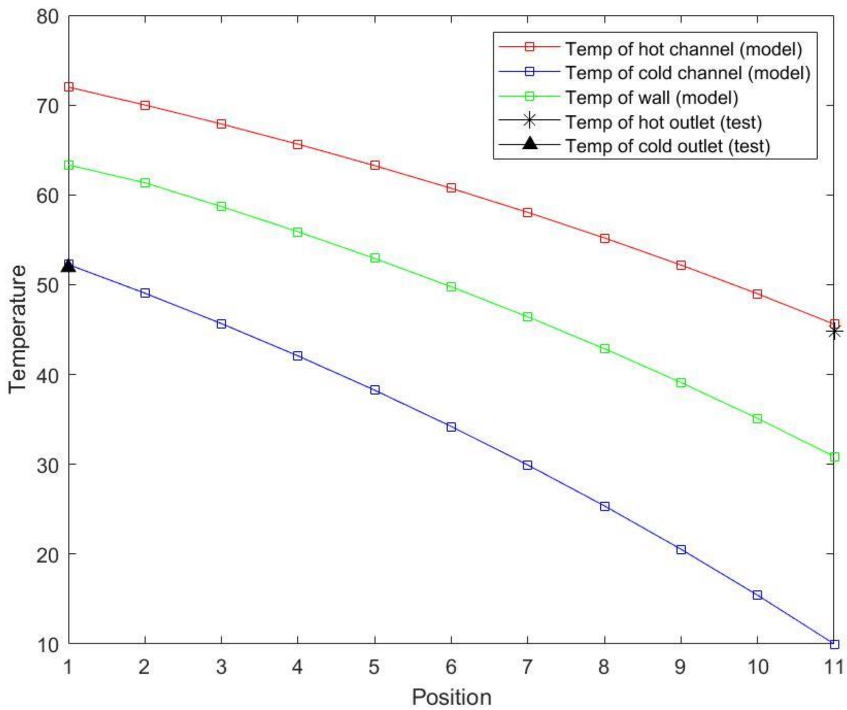


Figure 3.13: The temperature distribution of hot, cold channels and wall by position for 32 plates PHE

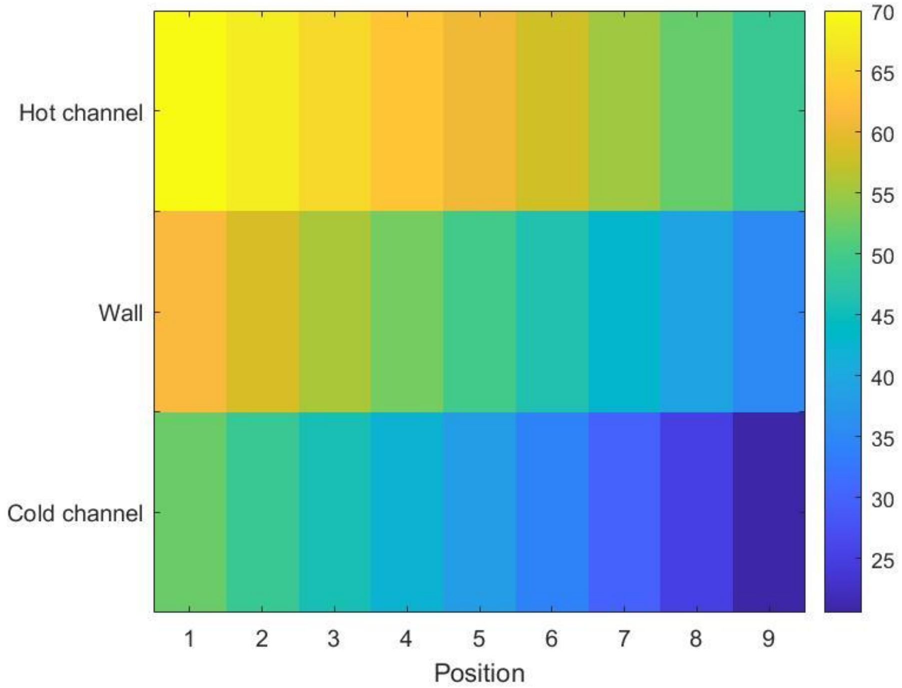


Figure 3.14: The temperature distribution in contour of hot, cold channels and wall by position for 32 plates PHE

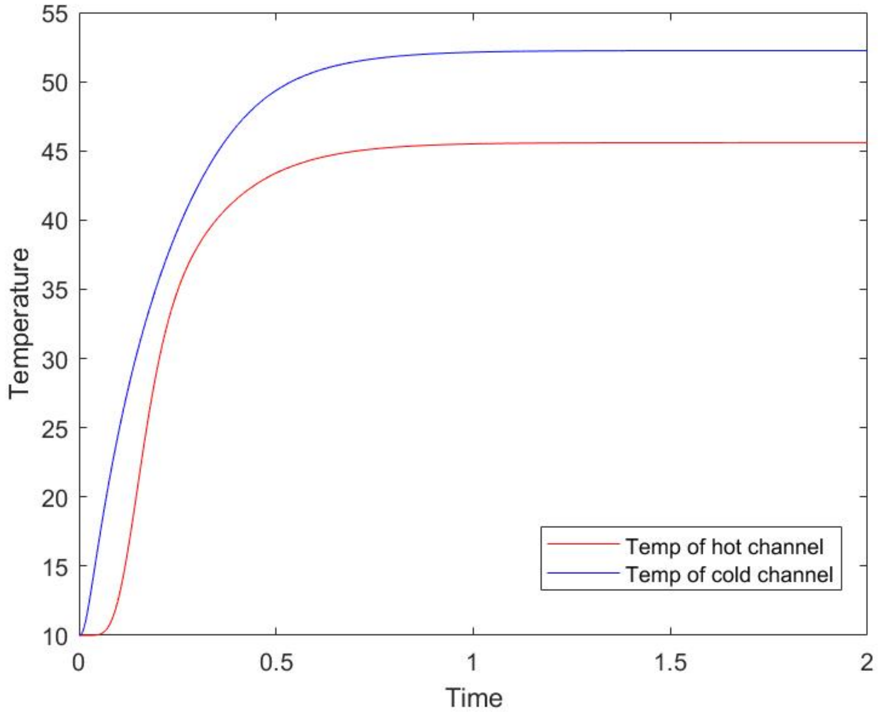


Figure 3.15: The temperature change of hot and cold channels by time for 32 plates PHE

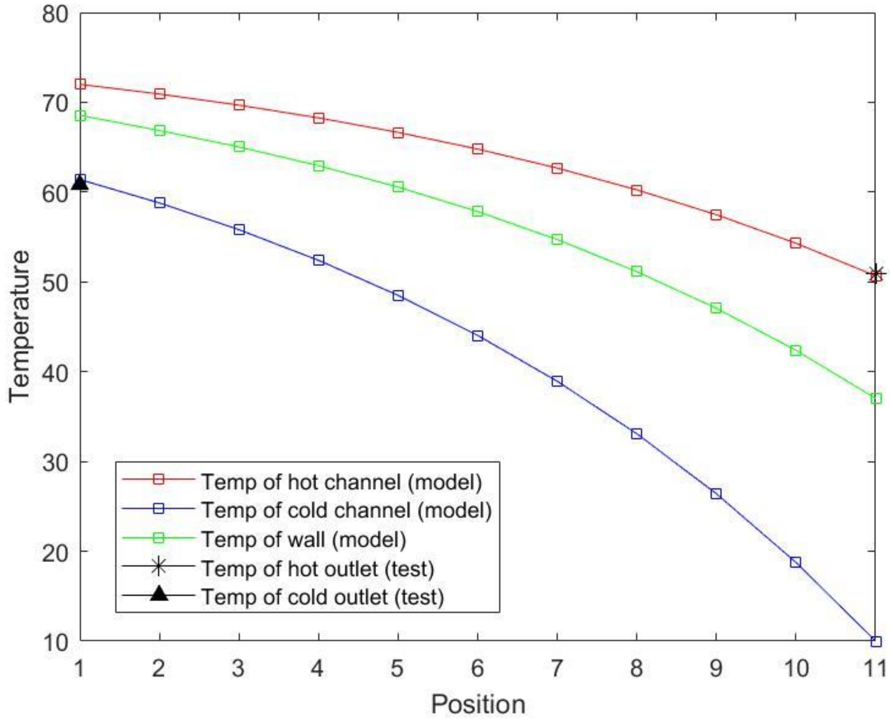


Figure 3.16: Temperature distribution of hot and cold channels by position for 30 plates PHE

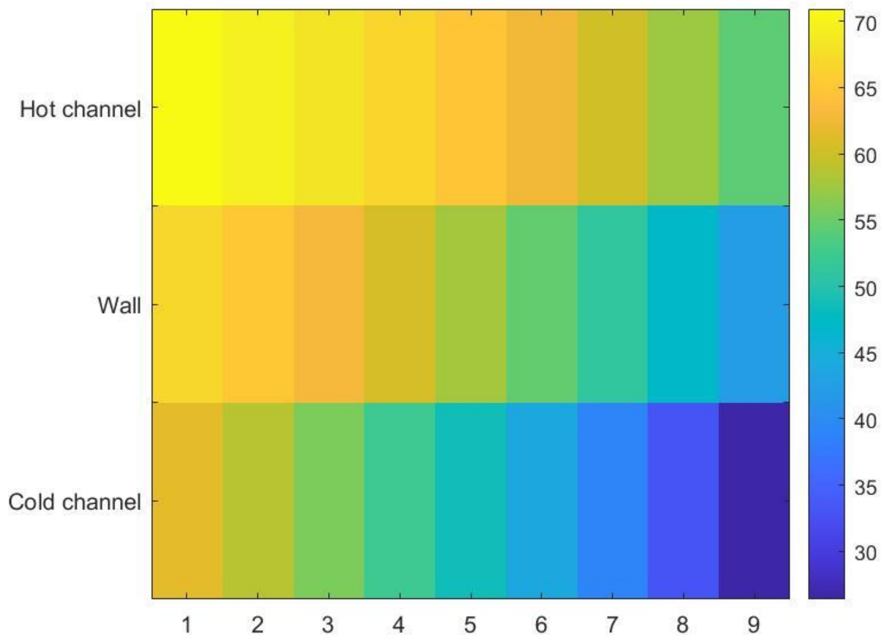


Figure 3.17: Temperature distribution in contour of hot and cold channels by position for 30 plates PHE

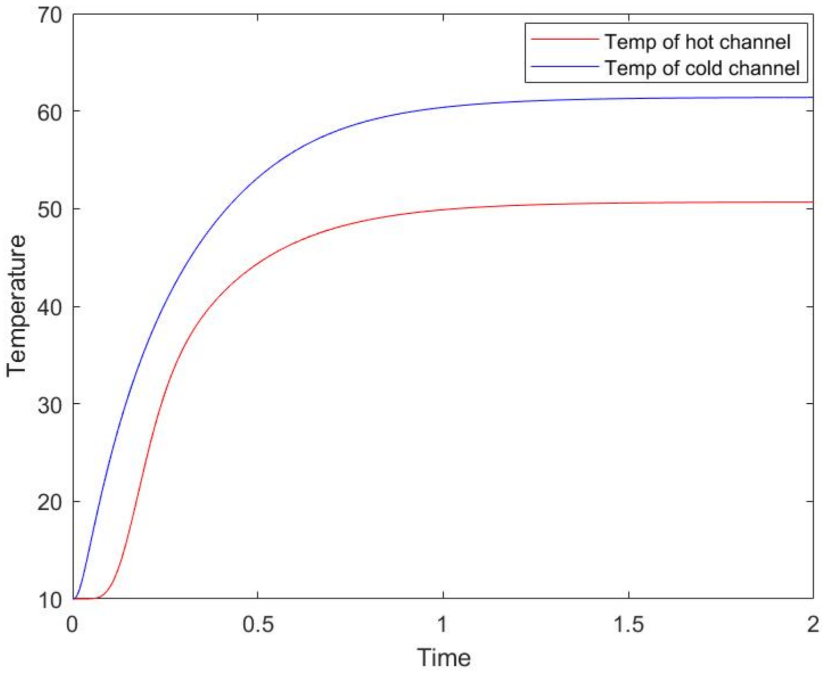


Figure 3.18: The temperature change of hot and cold channels by time for 30 plates PHE

In the results, the stabilization of the outlet temperatures is achieved less than 2 seconds. Therefore, the final time is selected as 2 seconds even though the cell size is selected as 10 as in the reference study. Considering the time step, flow and number of cells, the compatibility of the fluid in a time step was taken into consideration and the time step was selected as 0.01 seconds.

As a second validation, the model results were compared with the test results. The outlet temperatures of the hot and cold channels from test results and model results are also given in Table 3.5 for 32 and 30 plate PHEs. The cold channel (DHW) temperature for both types PHEs is closer than the hot channel (CH) temperatures for 32 plates of PHE, in contrast to the results of 30 plates of PHE. The errors are lower than 2%, therefore they are considered acceptable. As a result of the second validation, the model is considered as correct.

Table 3.5: The comparison of model and test results

	32 plate PHE			30 plate PHE		
	Model Results	Test Results	Error (%)	Model Results	Test Results	Error (%)
Hot channel outlet temperature	45.6 °C	44.8 °C	1.78	50.6 °C	50.9 °C	0.58
Cold channel outlet temperature	52.2 °C	51.9 °C	0.57	61.4 °C	60.8 °C	0.97

In the second part of the modeling study, the fouling resistance coefficients are used. The calculated fouling resistance coefficients from experiments are used in the model as shown in Equation (2.22) and (2.23). The main goal is validating the model accuracy for PHEs when the fouling resistance coefficients are taken into consideration. The expected result is that model should give the similar results to the experimental result based on the change in outlet temperatures caused by fouling. The model and test results are given in Figure 3.19 and 3.20 for 32 and 30 plates PHEs, respectively.

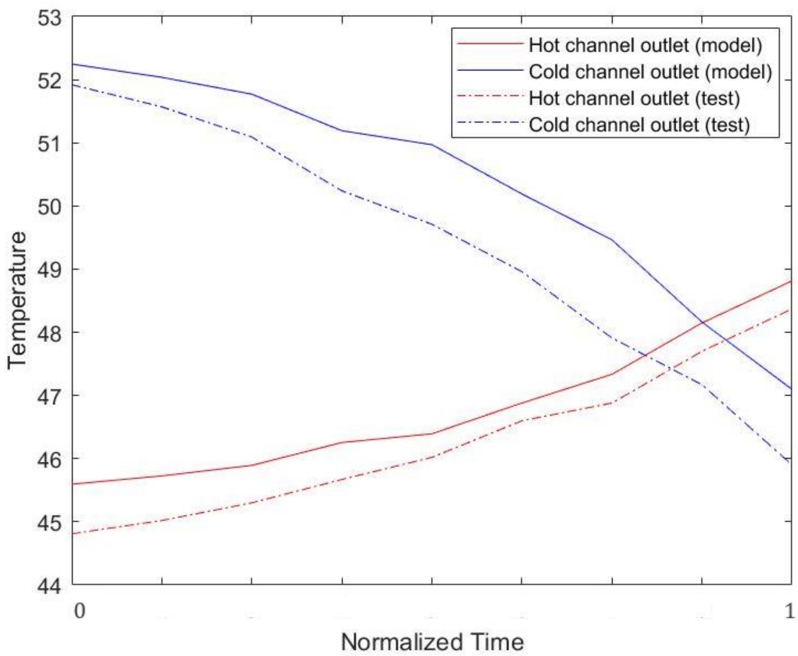


Figure 3.19: Comparison of model and test results based on fouling resistance coefficients for 32 plates PHE

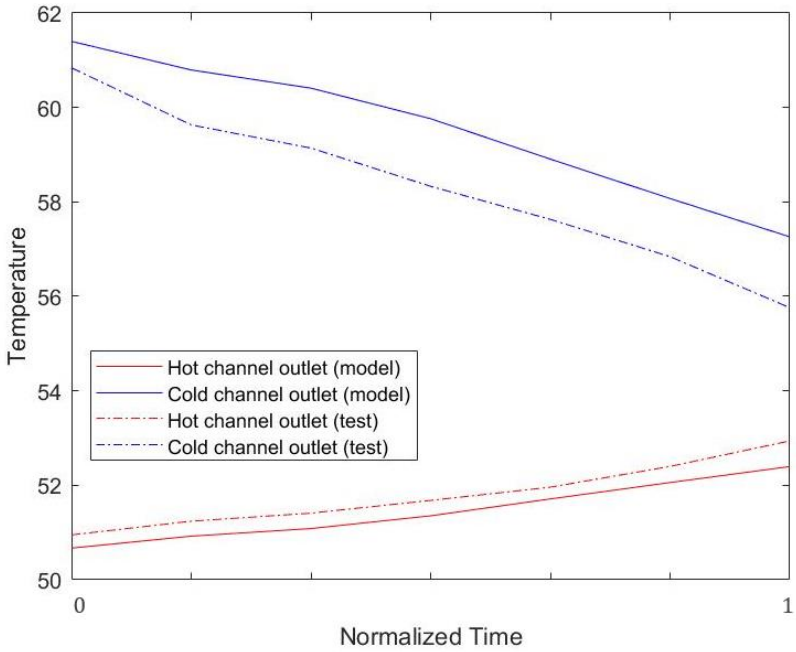


Figure 3.20: Comparison of model and test results based on fouling resistance coefficients for 30 plates PHE

In the results of the comparison of the model and test results based on fouling resistance coefficients, normalized time is considered as before when the experimental results are given. As in the comparison of the model and test results for faulty conditions, the hot channel (CH) temperatures are closer than the cold channel (DHW) temperatures. For 32 plate PHE, the hot channel model results are significantly close to the testing results. The difference in the worst case is lower than 2°C. This tolerance is considered acceptable and enough to evaluate that the model is running correctly.

The cold outlet temperature is the determining factor for customer comfort in combi-boiler appliances. The cold outlet temperature change during fouling for 32 and 30 plates PHE by normalized time is given in Figures 3.21 and 3.22. The indicated reference value of cold outlet temperature is the one that is representing the ideal case, where no fouling is seen. The cold outlet temperature should be within the tolerance of 1% as indicated in the figures with the reference value in case there is no fouling.

The created model provides a calculation of the required CH (hot channel) temperature to achieve the reference cold channel outlet temperature. When the maximum fouling case that was evaluated (Zone8) is considered, the required hot channel inlet temperatures to obtain the required cold channel outlet temperature by the customer as a setpoint are given in Figure 3.23 and 3.24.

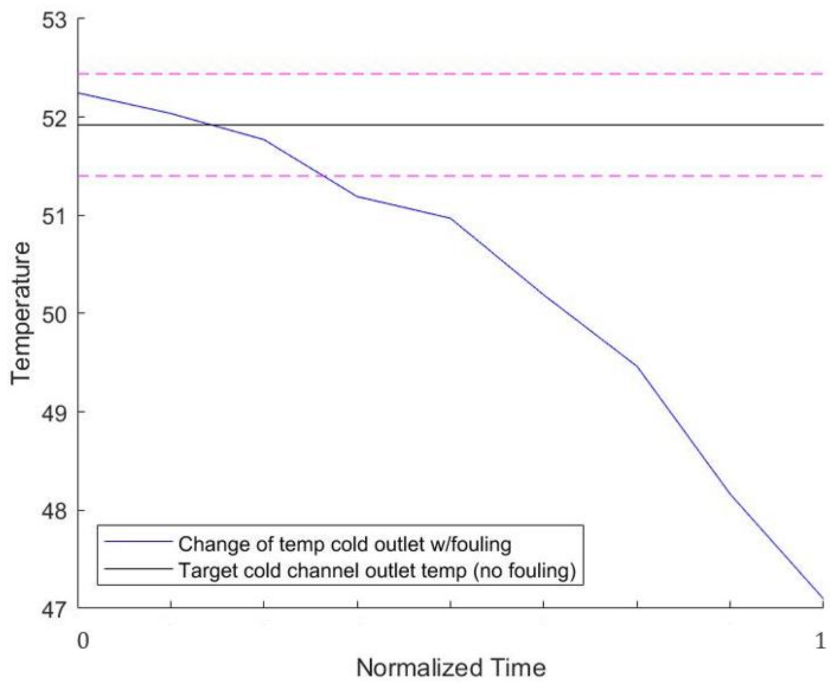


Figure 3.21: Change of cold channel outlet temperature during fouling for 32 plates PHE

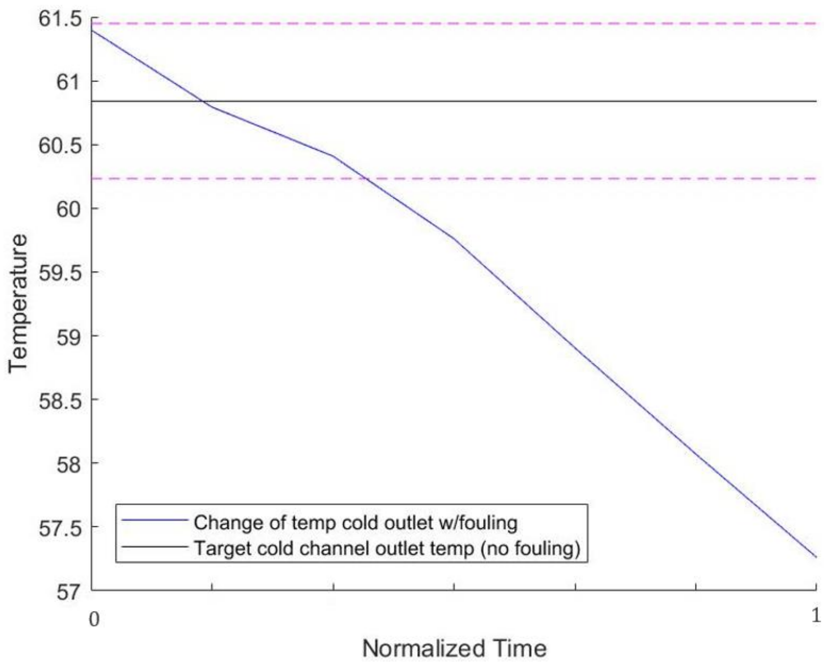


Figure 3.22: Change of cold channel outlet temperature during fouling for 30 plates PHE

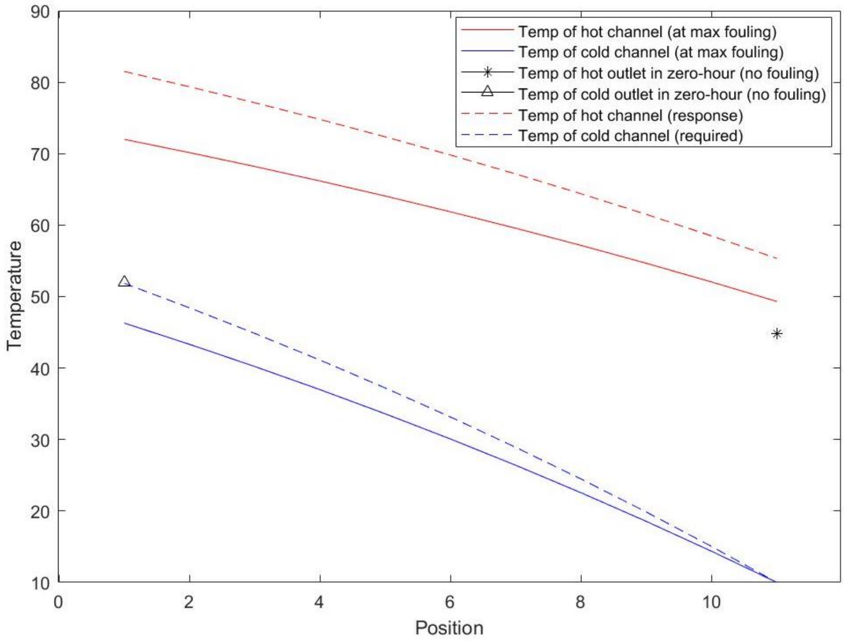


Figure 3.23: Temperature distribution of the response of the hot channel to reach the required ideal cold channel temperature in case of maximum fouling for 32 plates PHE

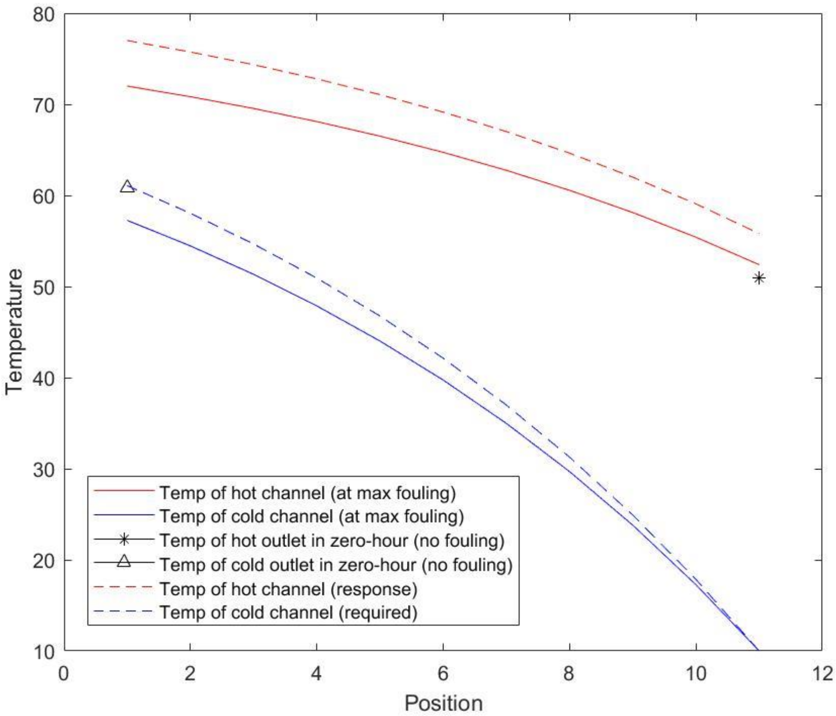


Figure 3.24: Temperature distribution of the response of the hot channel to reach the required ideal cold channel temperature in case of maximum fouling for 30 plates PHE

The existence of the fouling layer on the plate surfaces causes a decrease in customer comfort. Reaching the required setpoint of DHW defined by the customer needs more energy when there is fouling. This leads to an increase in energy consumption, i.e., natural gas consumption. The required additional power to reach the setpoint when there is maximum fouling is calculated.

The setpoints for 32 and 30 plates are selected from the test outputs (Table 3.4). As shown in Figures 3.23 and 3.24, CH inlet temperature should be 81 and 77 °C to reach the setpoints that are 51.9 and 60.8 °C at maximum fouling case. The appliance should heat up the water in the heat cell (primary heat exchanger) up to these temperatures to reach the required setpoint temperature in DHW. The required additional power is calculated for 32 and 30 plate PHEs. The maximum required additional powers are shown in Table 3.6. The power changes by fouling, i.e., fouling zones, are shown in Figure 3.25 and 3.26 for 32 and 30 plates PHEs, respectively.

Table 3.6: The comparison of model and test results

	Assumed setpoint of DHW outlet temperature (required) (°C)	DHW outlet temperature at maximum fouling (model results) (°C)	CH inlet temperature in case no fouling (°C)	CH inlet temperature to reach the setpoint at maximum fouling (response) (°C)	Additional required power to reach the setpoint at maximum fouling (kW)
32 Plates PHE	51.9	47.1	72	81	16
30 Plates PHE	60.8	57.3	72	77	7

The heat output when there is no fouling is approximately 50 kW for 32 and 35 kW for 30 plates PHE. With the additional required power, the heat outputs to reach the setpoint in case of maximum fouling are 66 and 42 kW for 32 and 30 plates PHEs, respectively. The amount of natural gas consumption is approximately 0.5 m^3/h for 50 kW appliances and 0.35 m^3/h for 35 kW appliances. The required natural gas consumption to reach the setpoint in case maximum fouling would be 0.66 and 0.42

m^3/h for 32 and 30 plates PHEs, respectively. The price of 0.1 m^3/h natural gas is 2.6₺ as of 01.10.2022 from Enerji Piyasaları İşletme A.Ş. (Energy Market Operation Inc.). Therefore, the additional costs are 83.2 and 36.4 ₺ for 32 and 30 plates PHEs, respectively.

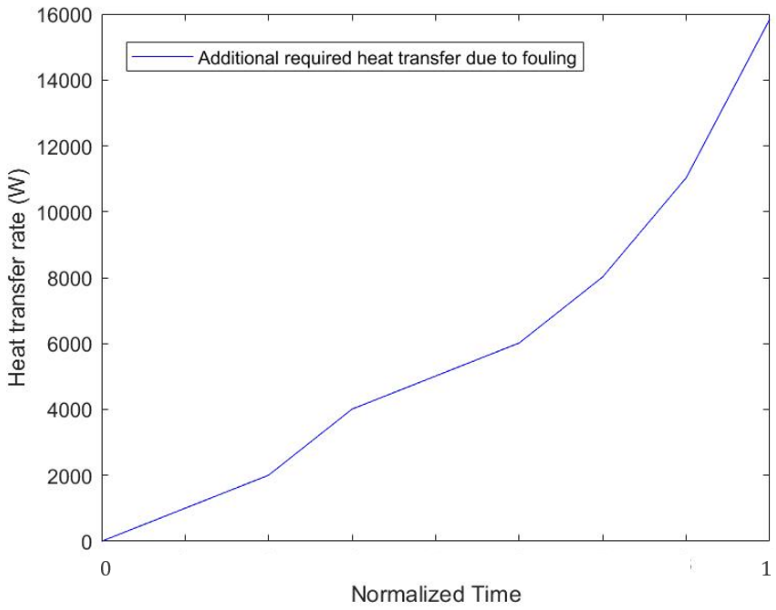


Figure 3.25: Change of additional required power to reach the setpoint by fouling in normalized time for 32 plates PHE

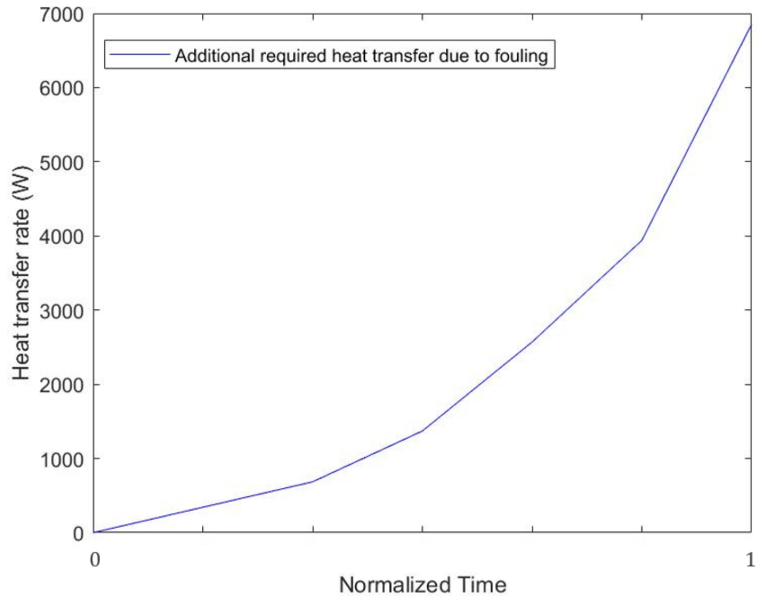


Figure 3.26: Change of additional required power to reach the setpoint by fouling in normalized time for 30 plates PHE

3.3 Algorithm Results

The three algorithms, Naïve Bayes, k-nearest neighbors and decision tree models are applied to each case listed in Table 2.2. In each case, the training set and testing set is selected with k-fold cross validation to avoid overfitting as mentioned in 2.2.2. The main results are calculated as accuracy. The data implementation and model training processes are evaluated in the Classification Learner App in MATLAB programming tool as shown in Figures 3.27 and 3.28. During data implementation, the features, i.e., predictors are in the type of double integer while the responses are in the type of categorized data. Thus, the application can distinguish the training predictors data and corresponding training classes as can be seen in Figure 3.27.

The classified training data for Case 1 is shown in a scatter plot in Figure 3.29, where the DHW outlet temperature is on x-axis and CH outlet temperature is on y-axis. The distribution of the training data shows the distinguishability of the training set.

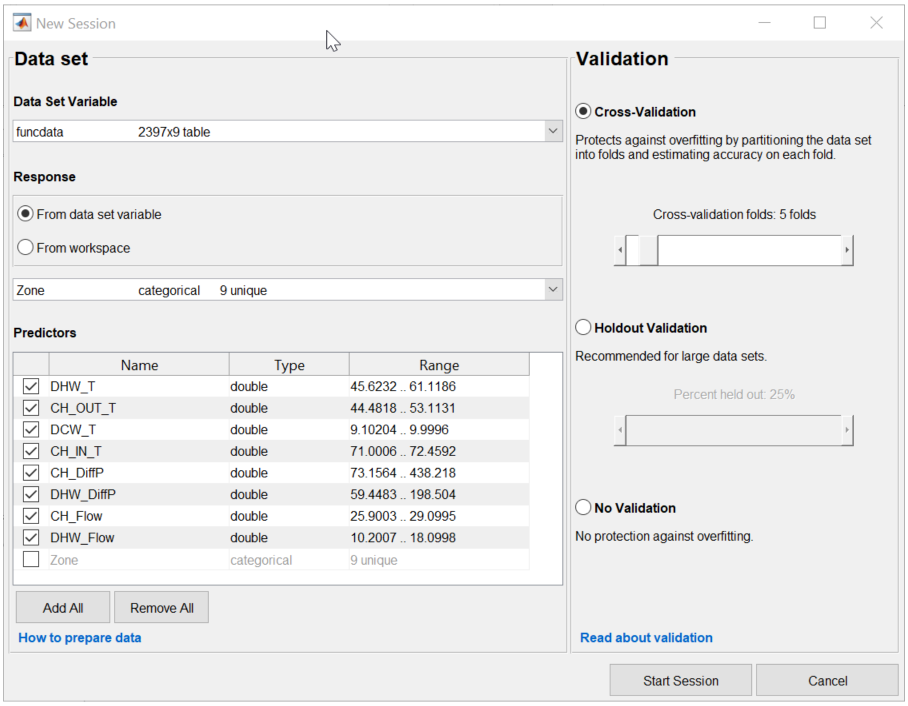


Figure 3.27: The data implementation in Classification Learning App in MATLAB.

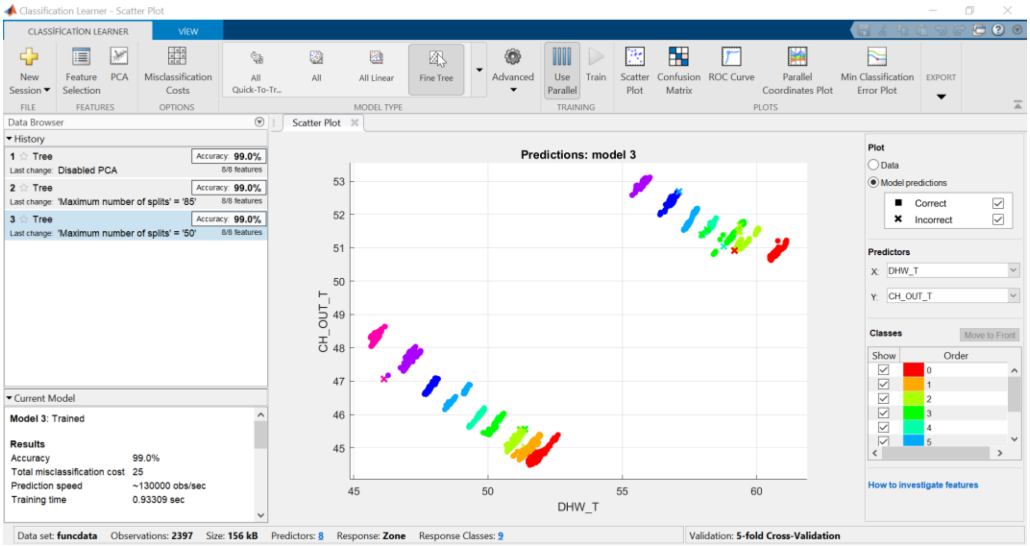


Figure 3.28: The model training process in Classification Learning App in MATLAB.

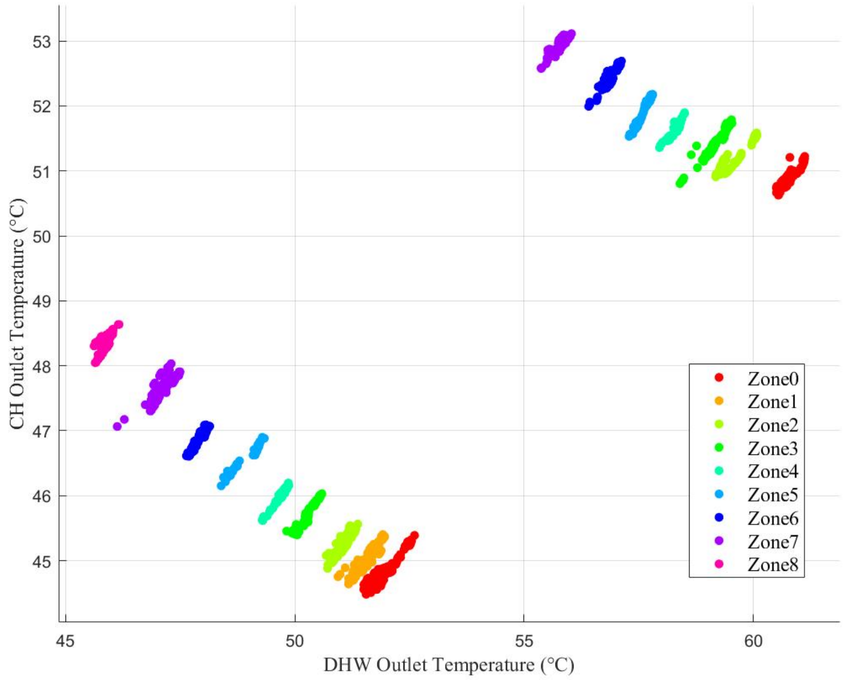


Figure 3.29: The training data distribution shown in DHW outlet temperature vs. CH outlet temperature of Case 1.

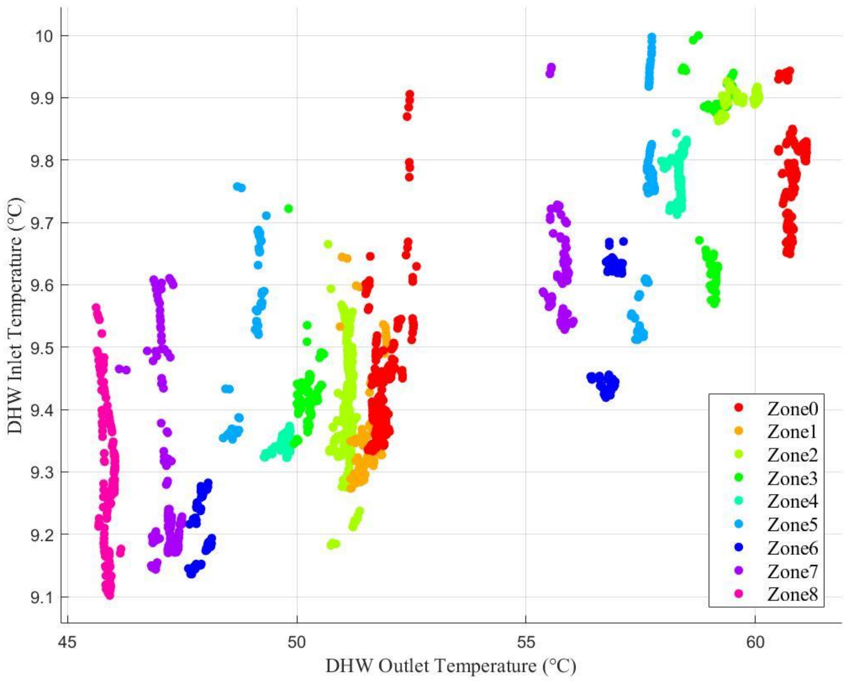


Figure 3.30: The training data distribution shown in DHW inlet vs. outlet temperature of Case 1.

In Figure 3.29, the data can be separated easily, in contrast to Figure 3.30, where the DHW inlet and outlet temperature distribution is shown. The zones, i.e., training classes are not very easily distinguishable when it is compared to Figure 3.29. This inference results in the generation of other cases, to see the difference between the case with selected all parameters and the case with selected fewer parameters.

For Case 1, three algorithms are applied with different features are shown with the accuracies in Table 3.7. For the decision tree model, Gini's Index is used as a split criterion and four different maximum number of splits are applied. As can be seen in Table 3.7, there is no difference between the accuracies even the maximum numbers of splits are changed.

Table 3.7: The accuracies of Case 1 according to model features

Models	Features	Accuracy (%)
Decision Tree (Split criterion=Gini's Index)	Maximum number of splits=100	99.0
	Maximum number of splits=85	99.0
	Maximum number of splits=50	99.0
	Maximum number of splits=25	99.0
Kernel Naïve Bayes (Kernel type=Triangle)	Support=Positive	99.9
	Support=Unbounded	99.9
k-nearest neighbors (Distance metric=Euclidean)	Number of neighbors=1	99.5
	Number of neighbors=10	96.4
	Number of neighbors=15	95.1
	Number of neighbors=50	84.6

The confusion matrix is used to show the detailed prediction accuracies of each class of a model. In Figure 3.31, the confusion matrix of the decision tree model for Case 1 is shown. Here, the TPR indicates the true positive rates which is the rate of true predicted data overall data and FNR indicates the false negative rates which is similarly the rate of false predicted data overall data. When the overall accuracy of the decision tree model for Case 1 is considered as 99.0%, the FNR values are expected to be this small. The highest FNR value is seen in the prediction of 3rd class, i.e., Zone 3. The model predicted the 2.9% data of Zone 3 as Zone 4. The model is successfully predicting the data from Zone 5 and Zone 8.

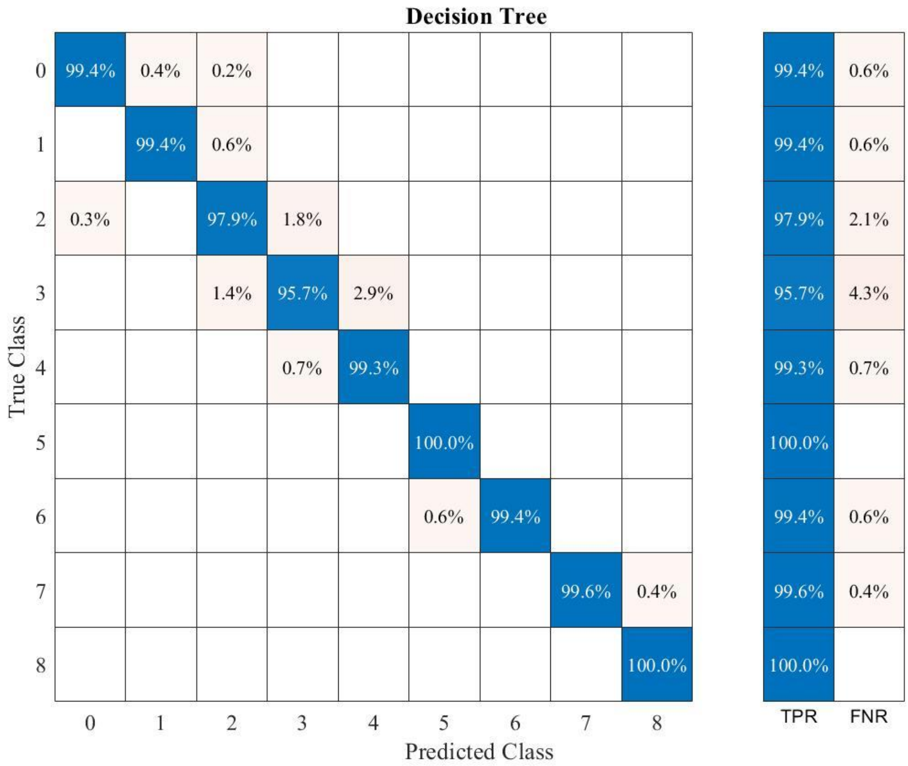


Figure 3.31: The confusion matrix of Decision Tree model for Case 1.

The classification tree with nodes and leaves is given in Figure 3.32 for the maximum number of splits is 100. The algorithm trains itself by selecting the pressure difference of DHW for the main node. This shows that pressure difference data gives more precise distinguishability than the other predictors.

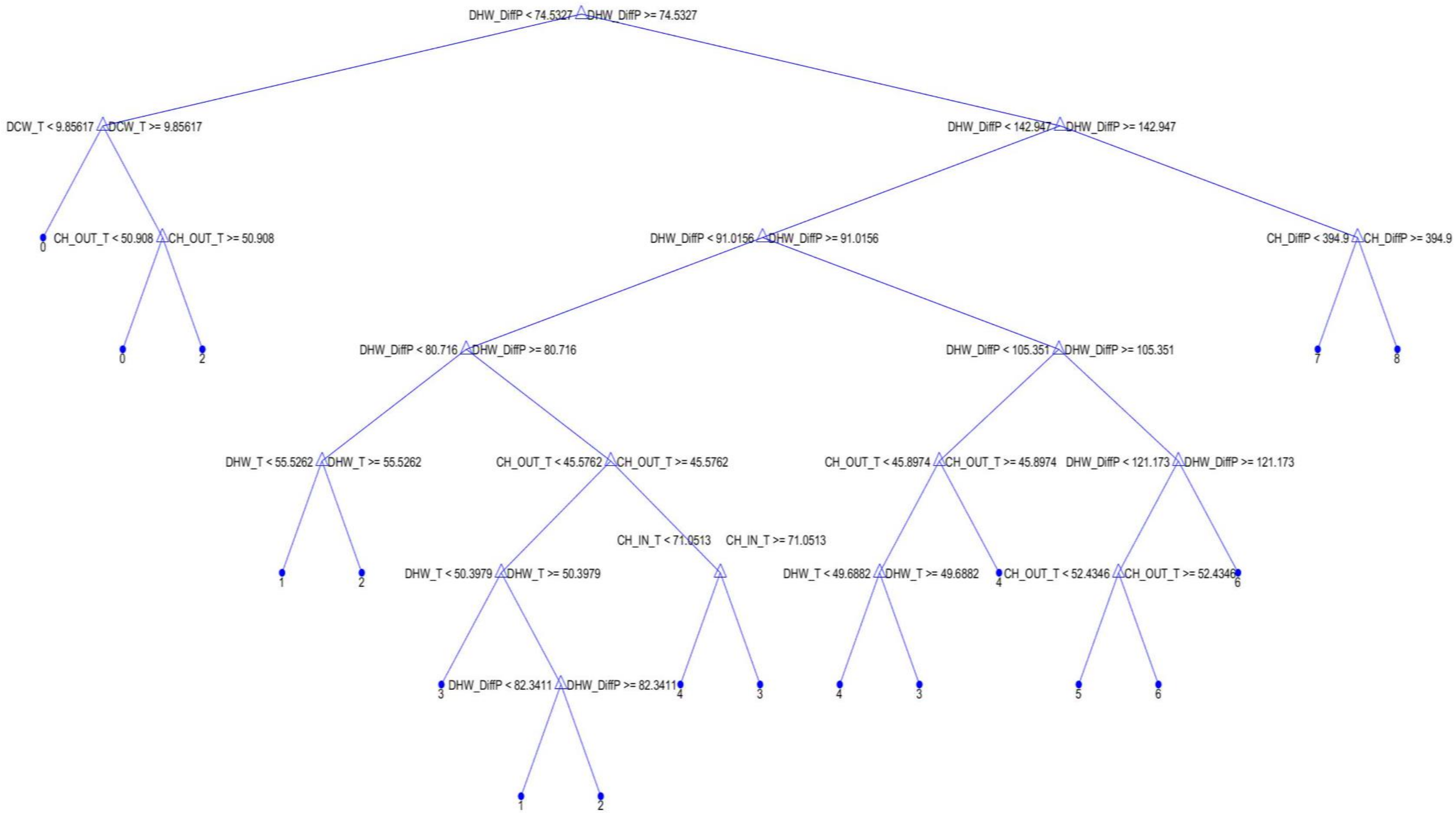


Figure 3.32: The classification tree designation of the decision tree model for Case 1.

For the Kernel Naïve Bayes model, the “triangle” kernel type is selected. The Naïve Bayes model is trained based on the independence of predictors. However, both the support functions, positive indicates dependency and unbounded indicates independency, are applied and it is seen that the accuracy is not changed. This shows that Naïve Bayes model is valid for this case whether the predictors are independent of their or not.

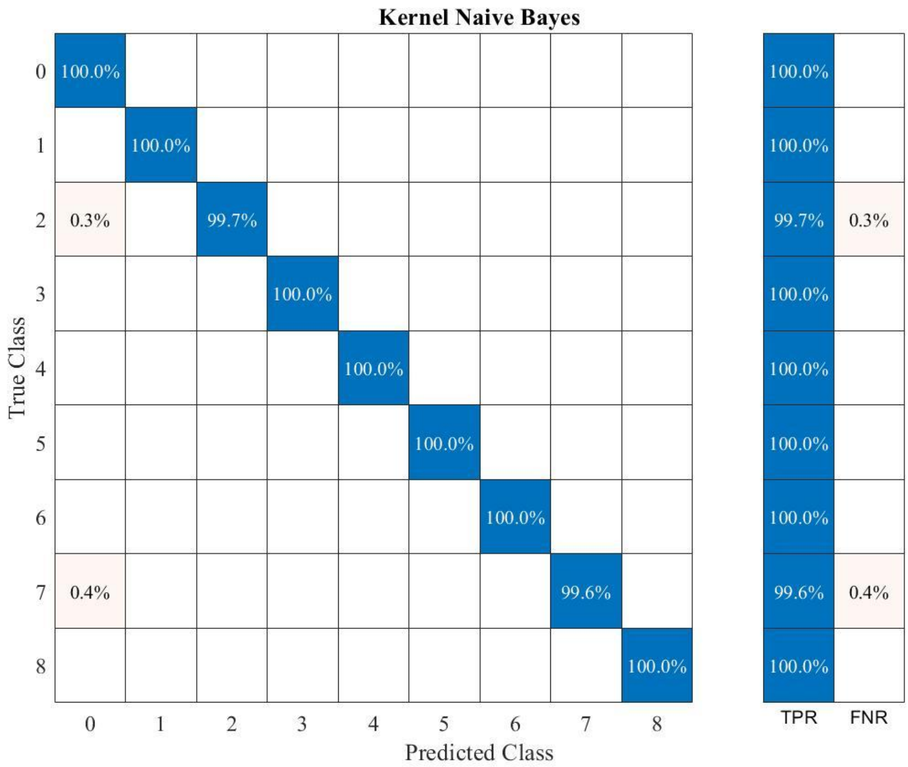


Figure 3.33: The confusion matrix of Kernel Naïve Bayes model for Case 1.

In Figure 3.33, the confusion matrix of Kernel Naïve Bayes model is given. The overall accuracy for both features is obtained as 99.9% which results in overfitting. The data is successfully classified by the model except for Zone 2 and 7.

The k-nearest neighbor model is applied with the various number of neighbors. The number of neighbors is selected as 1, 10, 15 and 50. The best accuracy is achieved while the number of neighbors is selected as 1. This results in when all parameters are used as predictors, the training data and test data would be similar and therefore the model is too much fitted to the training data, yet it is achieved to the highest accuracy. The model that has the lowest accuracy with the number of neighbors 50, is represented in the confusion matrix in Figure 3.34. The highest FNR value is seen at Zone 3. The 53.4% of data that classified as Zone 3 is predicted as Zone 2. The false prediction of data classified as Zone 3 is also seen in decision tree algorithm. When the data distribution in Figure 3.29 is considered, Zone 3 and Zone 2 have similar ranges, they are hard to be distinguished. Thus, the highest FNR values are seen between these zones.

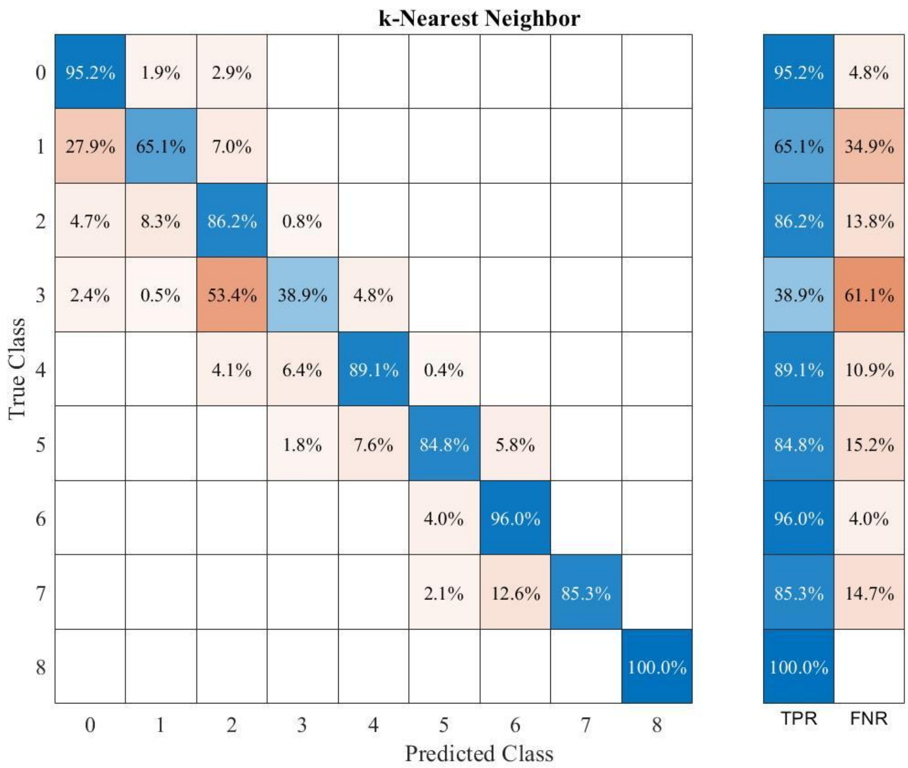


Figure 3.34: The confusion matrix of k-Nearest neighbor model for Case 1.

For Case 2, as it is listed in Table 2.2, the overall heat transfer coefficient is used as a predictor, whereas the responses are kept the same as in Case 1, from Zone 0 to 8. The same features that are used for the models in Case 1 are applied to Case 2. In contrast with the achieved accuracies in the Decision Tree model for Case 1, the obtained accuracies of the decision tree model for case 2 are changing with a maximum number of splits. The highest accuracy is achieved for the decision tree model with a maximum number of splits of 25. Additionally, the lowest accuracy is seen for the Decision Tree Model with a maximum number of splits between 100 and 85.

Table 3.8: The accuracies of Case 2 according to model features

Models	Features	Accuracy (%)
Decision Tree (Split criterion=Gini’s Index)	Maximum number of splits=100	94.7
	Maximum number of splits=85	94.7
	Maximum number of splits=50	94.4
	Maximum number of splits=25	95.8
Kernel Naïve Bayes (Kernel type=Triangle)	Support=Positive	95.7
	Support=Unbounded	95.7
k-nearest neighbors (Distance metric=Euclidean)	Number of neighbors=1	94.1
	Number of neighbors=10	95.2
	Number of neighbors=15	95.3
	Number of neighbors=50	95.9

The confusion matrix of the Decision Tree model with a maximum number of splits of 100 is given in Figure 3.35. The highest FNR value is encountered in Zone 6. The 30.9% of data classified in Zone 6 is predicted as Zone 0. The accuracies of the decision tree model for case 2 are higher than for case 1. The classification tree of the model is given in Figure 3.36. Zone 0 distinguished from the others for 32 plates PHE as can be seen in the first node of the tree. However, the Zone 0 classification is proceeded through the below nodes. The reason for that is the overall heat transfer coefficients of 30 plates PHE that are classified as Zone 0 are similar to the overall heat transfer coefficients of 32 plates PHE that are classified as Zone 6. This similarity in the data results in the observation of the highest FNR values between Zone 0 and Zone 6.

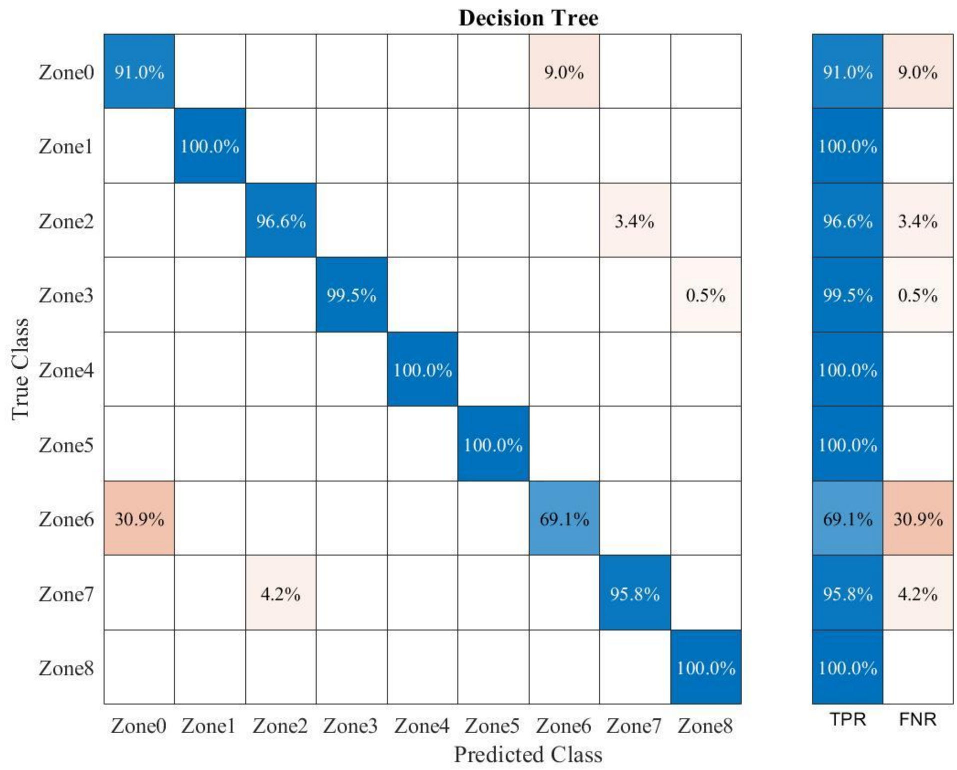


Figure 3.35: The confusion matrix of the Decision Tree model for Case 2.

The Kernel Naïve Bayes model gives the same result of accuracy with both features, as in Case 1. Whether the supporting function is selected to depict the independency or dependency, the accuracy is not changed, 95.7%. However, when the predictor is changed to the overall heat transfer coefficient, the accuracy decreases for Kernel Naïve Bayes model. The confusion matrix of Kernel Naïve Bayes model is given in Figure 3.37. Due to the closeness of the overall heat transfer coefficients of Zone 0 for 30 plates PHE and Zone 6 for 32 plates PHE, the highest FNR value is also observed at this point. The 44.6% of data that is classified as Zone 6 is predicted by the model as Zone 6, whereas the 55.4% of data is predicted correctly.

The k-nearest neighbor model is trained with the various number of neighbors, e.g., 1, 10, 15 and 50. While the number of neighbors is increasing the obtained accuracy of the model is increasing too. The predictor data, the overall heat transfer coefficients, have more distinguishability characteristic than all parameters that are applied in Case

1. This can be an inference that the lowest accuracy is observed with 1 number of neighbors. When the k, the number of neighbors, is selected as 1, the model tends to

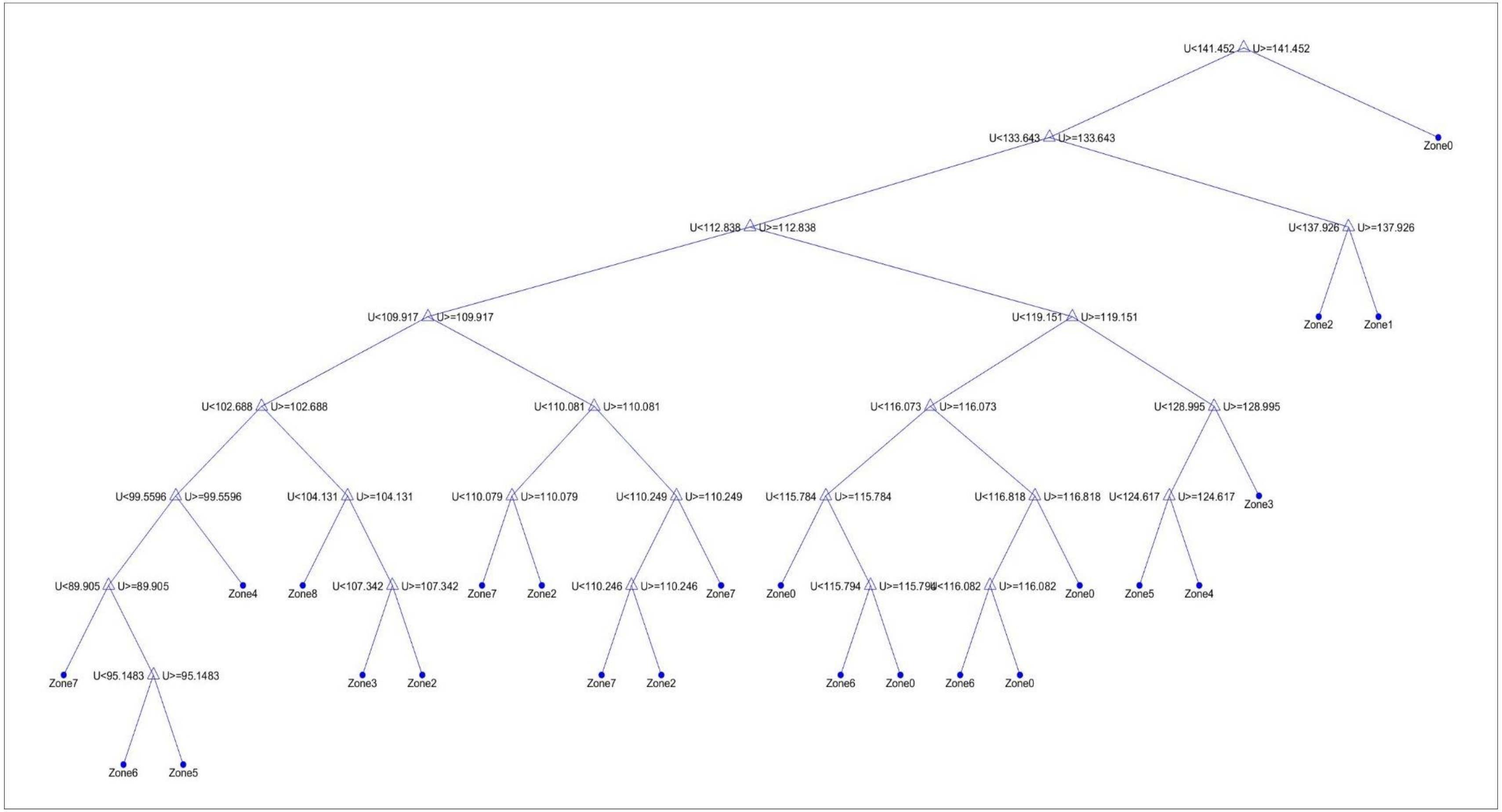


Figure 3.36: The classification tree designation of the decision tree model for Case 2.

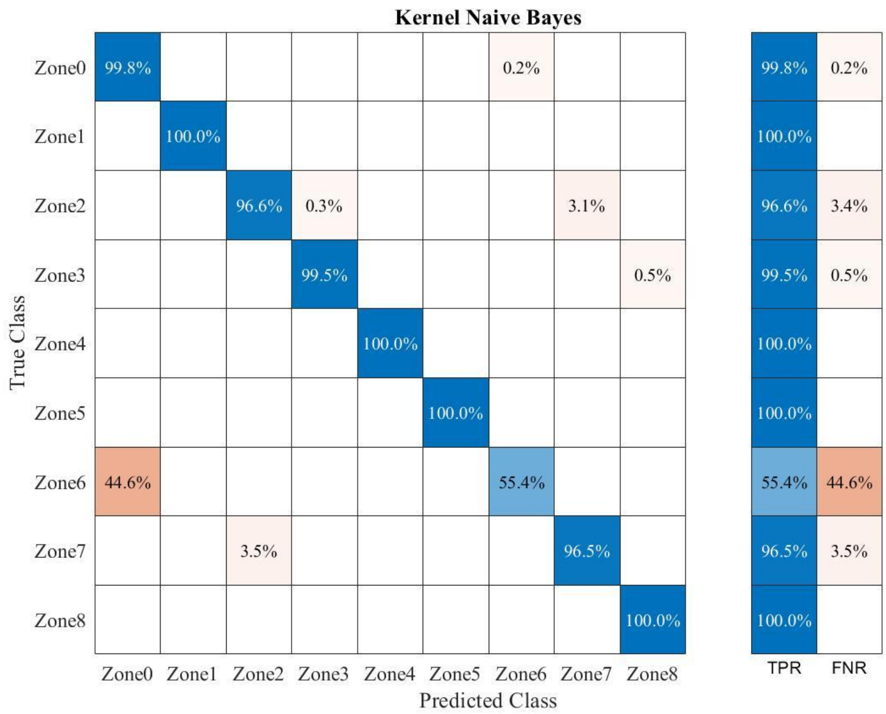


Figure 3.37: The confusion matrix of Kernel Naïve Bayes model for Case 2

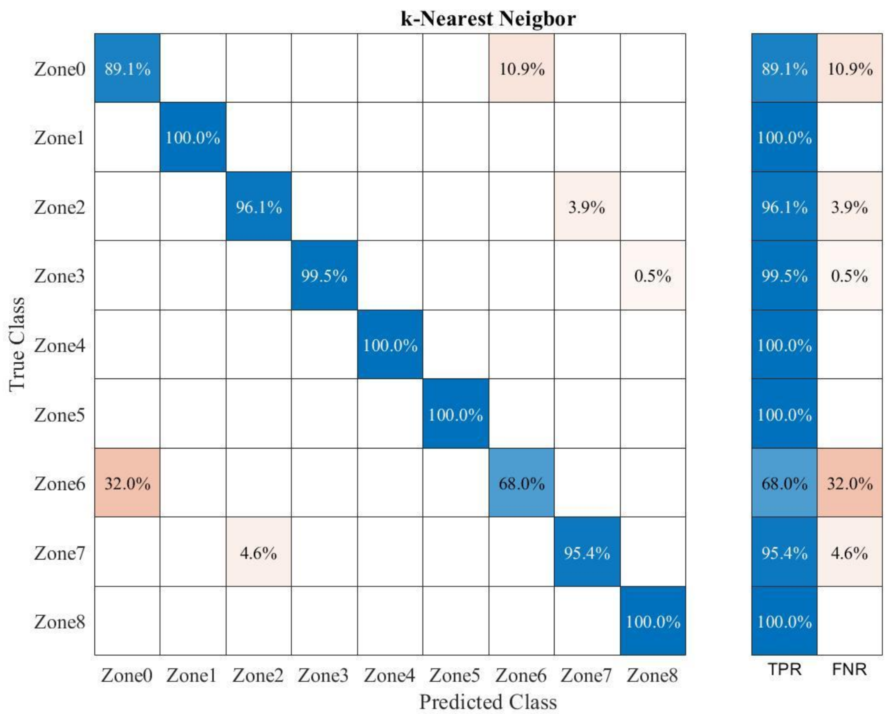


Figure 3.38: The confusion matrix of k-Nearest neighbor model for Case 2.

overfit the training set. Even the cross-validation method that is used, is preventing overfitting, if the test data is too similar to training data, in this case, if there is not any variance within the data, the model still can be overfitted. It is known that if the k is chosen as 1, this is the maximum fitting to the training data, i.e., indicates low bias. Even if the k number is 1, the lowest accuracy is observed. Thus, the inference can be the outcome that the variance within the overall heat transfer data is higher than the predictors of case 1. The highest accuracy observed for k number is 50. This shows that the k-nearest neighbor model achieved the required high bias- low variance characteristic.

Similarly, with the other models, the Zone 0 and Zone 6 distinguishability problem is observed in k-nearest neighbor model predictions. 32.0% of the data classified as Zone 6 is predicted as Zone 0. The FNR value of the k-nearest neighbor model is lower than the Kernel Naïve Bayes but slightly higher than the decision tree model.

For Case 3, the overall heat transfer coefficients are used as predictors as in Case 2, with a difference in response groups. The responses are selected in Case 3 as grouping Zone 0 to 5 by 2, and Zone 6,7 and 8 as one group, as shown in Table 2.2. The only difference from Case 2 is the response groups. The accuracies of the models that are trained in Cases 1 and 2 are shown in Table 3.9.

The decision tree model gives similar results to Case 2 even though the responses are different. Like in Case 2, the highest accuracy of the decision tree model is achieved with a maximum number of splits of 25. In addition, the accuracy variance within the decision tree model variations is small like Case 2. Thus, the decision tree model accuracies are merely dependent on the responses.

The confusion matrix of the decision tree model with a maximum number of splits of 50 is given in Figure 3.39. The FNR values are not larger than the ones that are observed in Case 2. The accuracies may be close, but the consistency of the model prediction is better for this grouping system method. The highest FNR value is obtained between Group 1 and 4. The 9.2% of data classified as Group 4 is predicted as Group 1, and the 1.8% of data classified as Group 4 is predicted as Group 2. This is added up to the total highest FNR value of 11.0%.

Table 3.9: The accuracies of Case 3 according to model features

Models	Features	Accuracy (%)
Decision Tree (Split criterion=Gini's Index)	Maximum number of splits=100	94.6
	Maximum number of splits=85	94.6
	Maximum number of splits=50	94.4
	Maximum number of splits=25	95.3
Kernel Naïve Bayes (Kernel type=Triangle)	Support=Positive	94.0
	Support=Unbounded	89.1
k-nearest neighbors (Distance metric=Euclidean)	Number of neighbors=1	94.4
	Number of neighbors=10	95.5
	Number of neighbors=15	95.4
	Number of neighbors=50	95.9

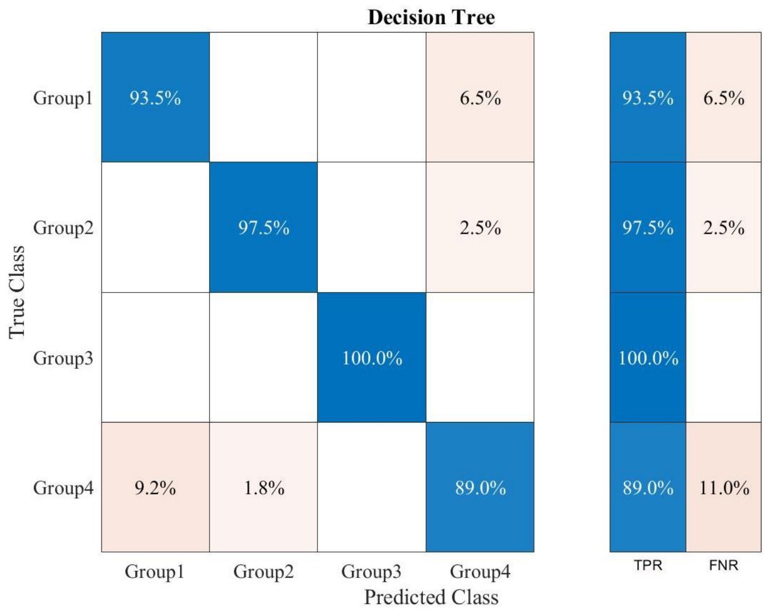


Figure 3.39: The confusion matrix of the decision tree model for Case 3.

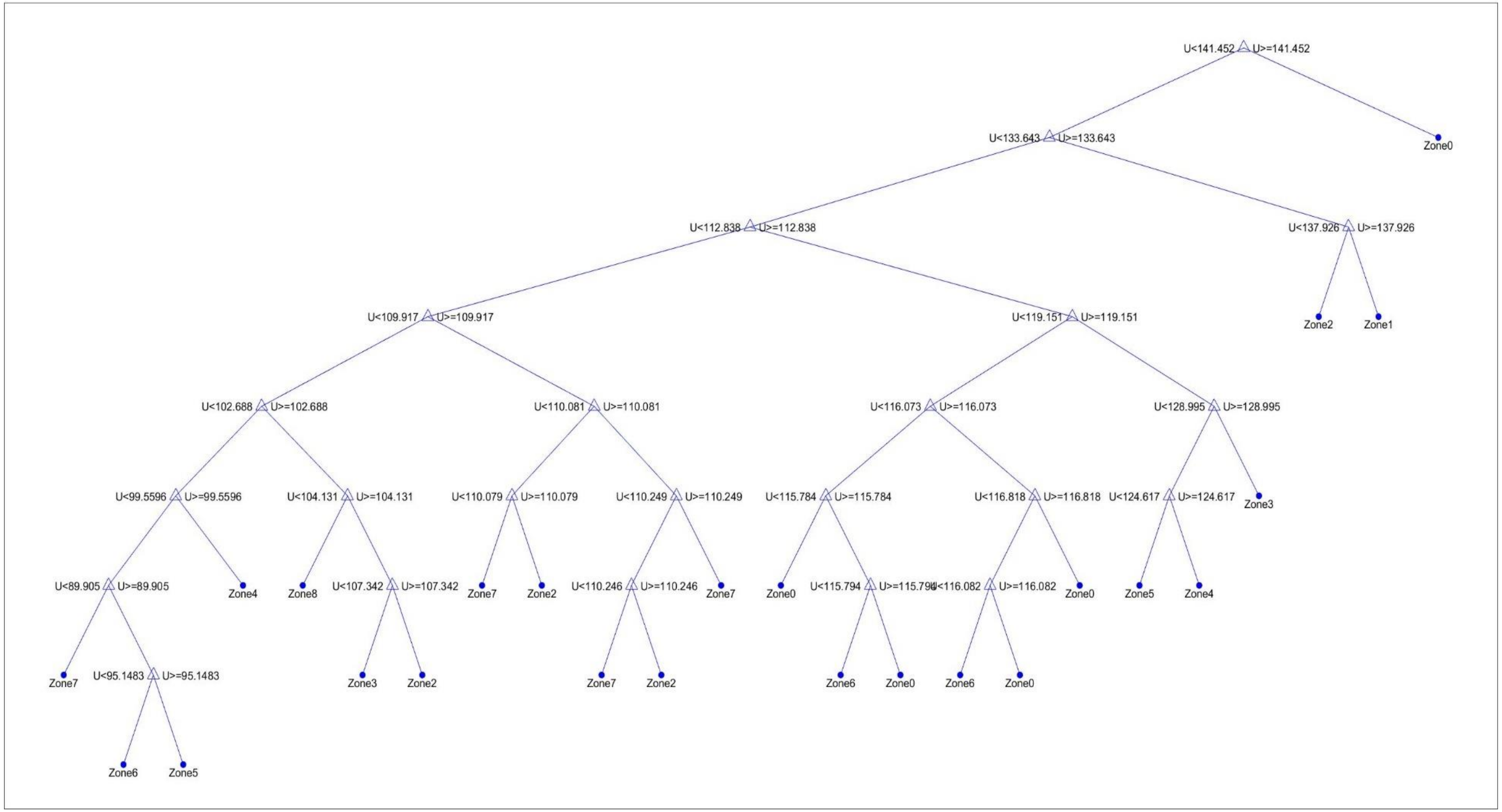


Figure 3.40: The classification tree designation of the decision tree model for Case 3.

In Figure 3.40, the classification tree of the decision tree model is represented. In similar with Case 2, Group 1 is generating the main node branch distinction, as can be seen in Figure 3.40. Through the below branches, it is observed that Group 1, and Group 4 classification branches are similar. This results in the model has difficulty distinguishing Group 1 and Group 4, similar to Zone 0 and Zone 6 as in Case 2.

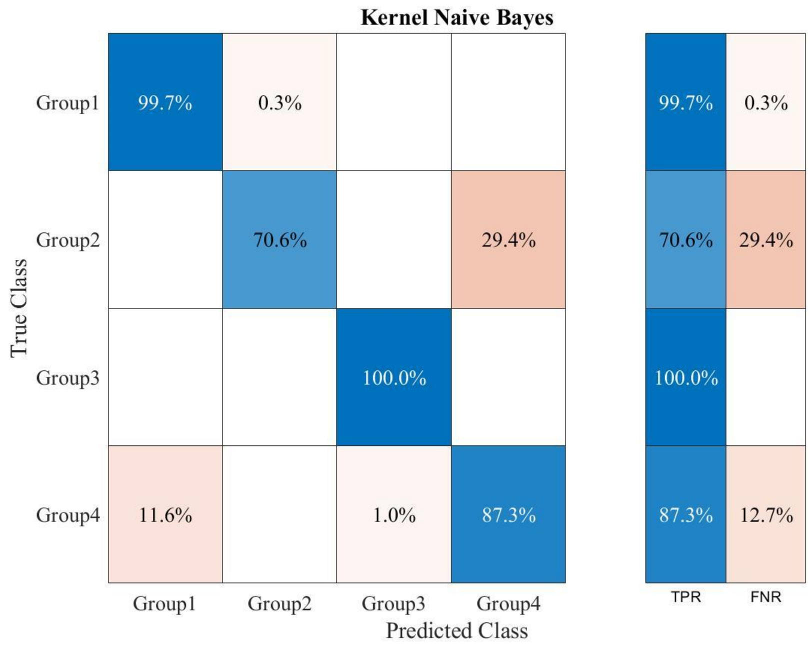


Figure 3.41: The confusion matrix of Kernel Naïve Bayes model for Case 3.

The Kernel Naïve Bayes model gives a similar result to Case 2 for the condition that the support function is applied as positive. However, when the support function is chosen as unbounded, the accuracy is considerably lower than Case 2.

The confusion matrix of the Kernel Naïve Bayes model with unbounded support function is given in Figure 3.41. The highest FNR value is obtained between Group 2 and Group 4, differently from the decision tree model. The 39.4% of data classified as Group2 is predicted as Group 4.

Similarly, the k-nearest neighbor model gives close results to Case 2. It can be deduced that the higher accuracy is achieved with a higher k number for Case 2 and 3. Like in Case 2, the k-nearest neighbor model is overfitted with the selection of the k number as 1. Thus, the lowest accuracy is observed for this variation of the k-nearest neighbor model while the model achieves the high bias and low variance characteristic with the selection of k number as 50 for Case 3, as in Case 2.

The confusion matrix of the k-nearest neighbor model for Case 3 with the k number as 15 is given in Figure 3.42. The highest FNR value is obtained between Group 1 and Group 4, similar to the decision tree model. The 10.4% of data classified as Group 4 is predicted as Group 1, and the 1.8% of data classified as Group 4 is predicted as Group 2, also similar to the decision tree model prediction.

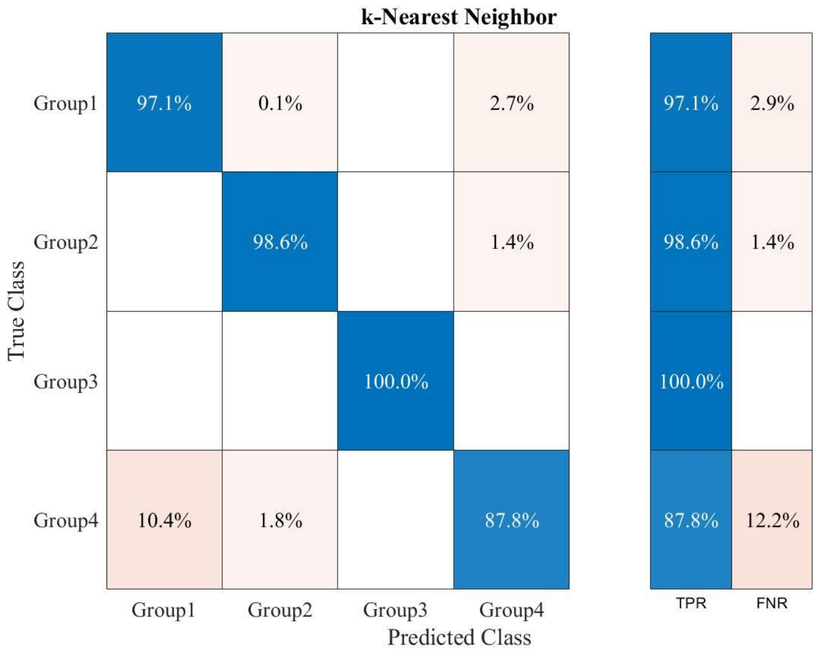


Figure 3.42: The confusion matrix of the k-nearest neighbor model for Case 3.

For Case 4, the same predictor, the overall heat transfer coefficient is applied to Case 2 and 3. In difference, the response group is selected as a representation of clogging percentages. The accuracies of the same models applied to the other cases are represented in Table 3.9.

The decision tree model again gives similar results to Case 2 and Case 3. The highest accuracy, 95.5% is achieved by selecting the maximum number of splits as 25 as it is in Case 2 and 3. This shows the decision tree model is considerably dependent on the predictor, not the responses.

Table 3.10: The accuracies of Case 4 according to model features

Models	Features	Accuracy (%)
Decision Tree (Split criterion=Gini's Index)	Maximum number of splits=100	94.8
	Maximum number of splits=85	94.8
	Maximum number of splits=50	94.7
	Maximum number of splits=25	95.5
Kernel Naïve Bayes (Kernel type=Triangle)	Support=Positive	95.0
	Support=Unbounded	89.7
k-nearest neighbors (Distance metric=Euclidean)	Number of neighbors=1	94.5
	Number of neighbors=10	95.3
	Number of neighbors=15	95.3
	Number of neighbors=50	95.7

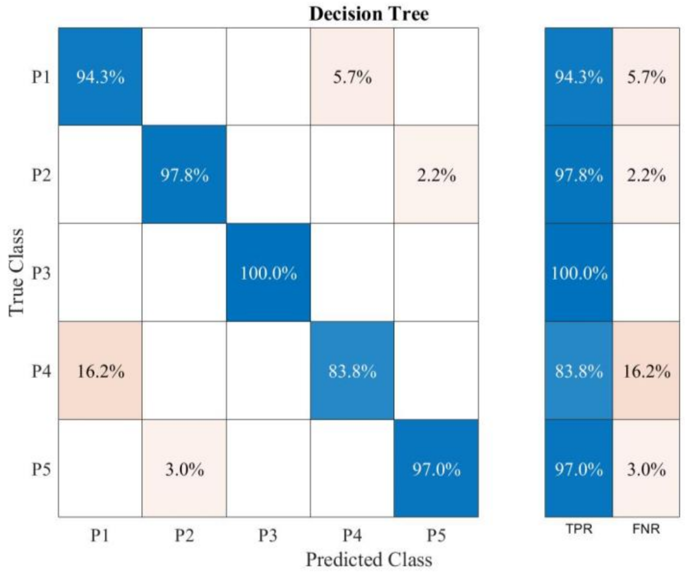


Figure 3.43: The confusion matrix of the decision tree model for Case 4.

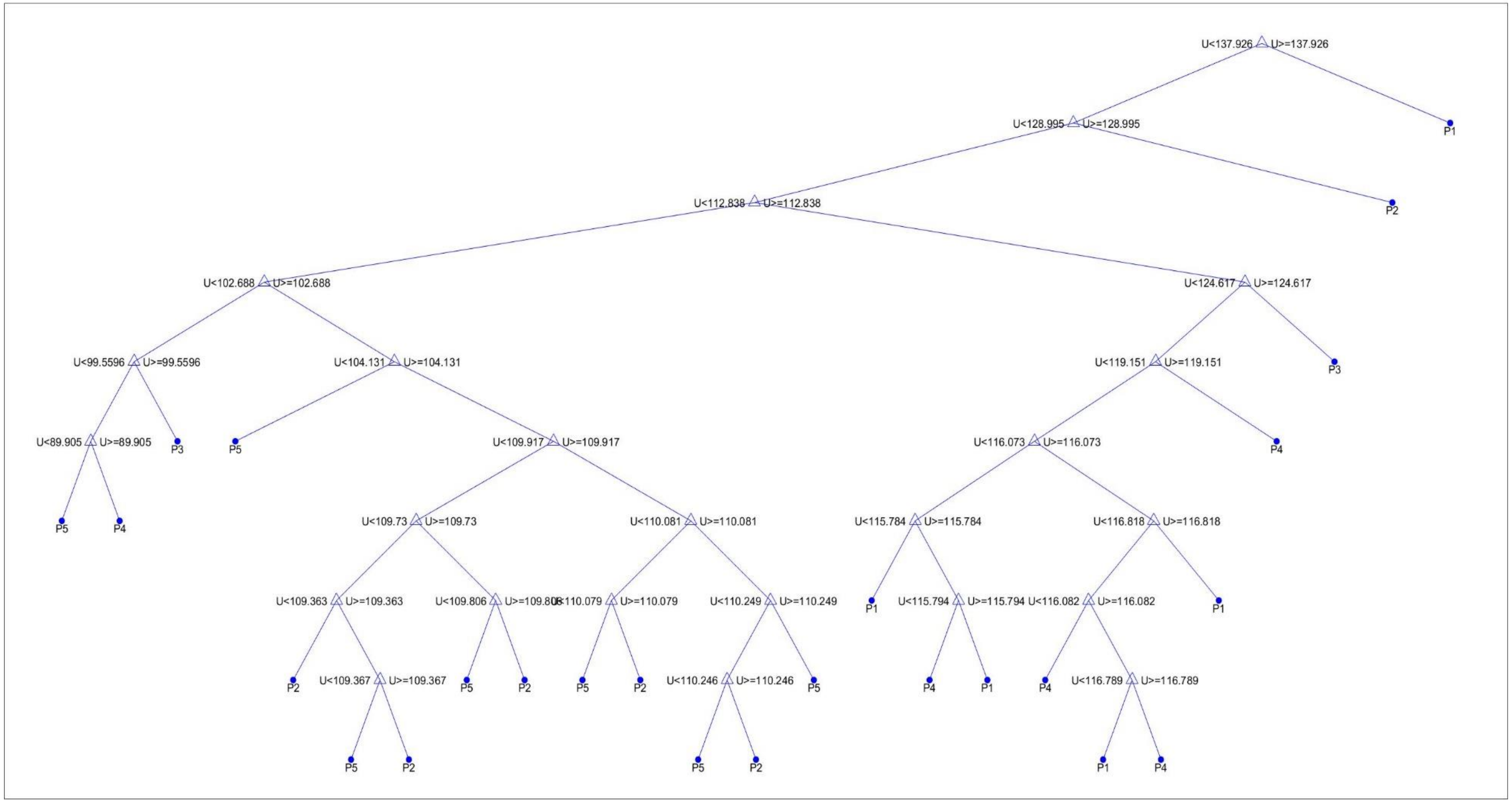


Figure 3.44: The classification tree designation of the decision tree model for Case 4.

In Figure 3.43, the confusion matrix of the decision tree model with the selection of a maximum number of splits of 100 is represented. The response groups are named P1 to P5. The highest FNR value is obtained between the P4 and P1 classes. The 16.2% of data classified as P4 is predicted as P1. The similarity between the overall heat transfer coefficients of 30 plates PHE classified in P1 with the overall heat transfer coefficients of 32 plates PHE classified in P4, results in this high false prediction rate, as in Case 2 and Case 3.

The classification tree of the decision tree model for Case 4 is given in Figure 3.44. Similar to Case 2 and 3, the P1 group generates the distinction of the main node. It is observed that it the difficulty to classify the P1 and P4, as represented in the confusion matrix, Figure 3.43.

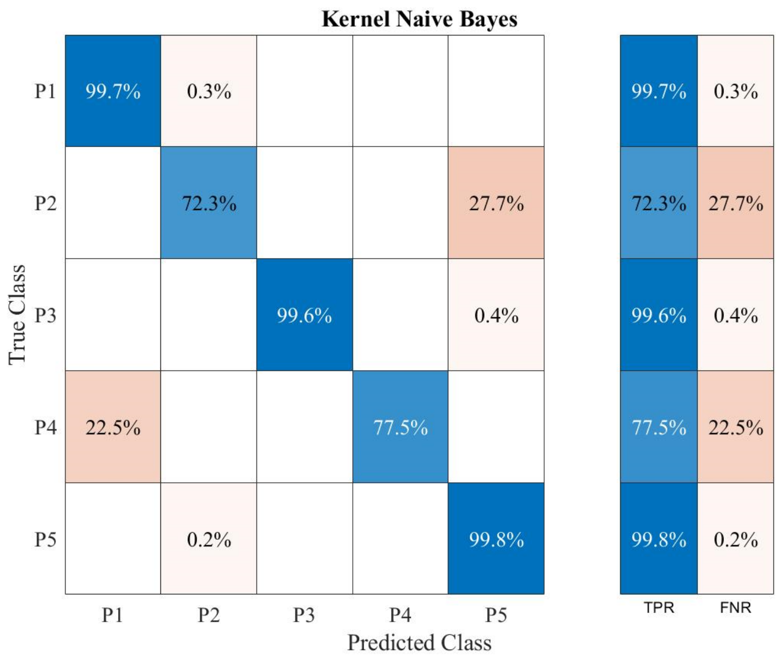


Figure 3.45: The confusion matrix of Kernel Naïve Bayes model for Case 4.

The Kernel Naïve Bayes model also gives similar result to Case 2 and 3. The predictors have more role in model predictions than the response groups. The highest accuracy, 95.0% is achieved by selecting the support function as positive. The model variation gives smaller results with the selection of unbounded support function, with 89.7% of accuracy.

In Figure 3.45, the confusion matrix of Kernel Naïve Bayes with the selection of an unbounded support function is represented. The highest FNR value is obtained between the P2 and P5. The pattern of the FNR value increase looks like similar to Case 2 and Case 3. The main reason is that even though the predictors are grouped in different responses, i.e., classified in training classes, the values of the predictors are the same with the Case 2 and 3. Thus, similar patterns are seen.

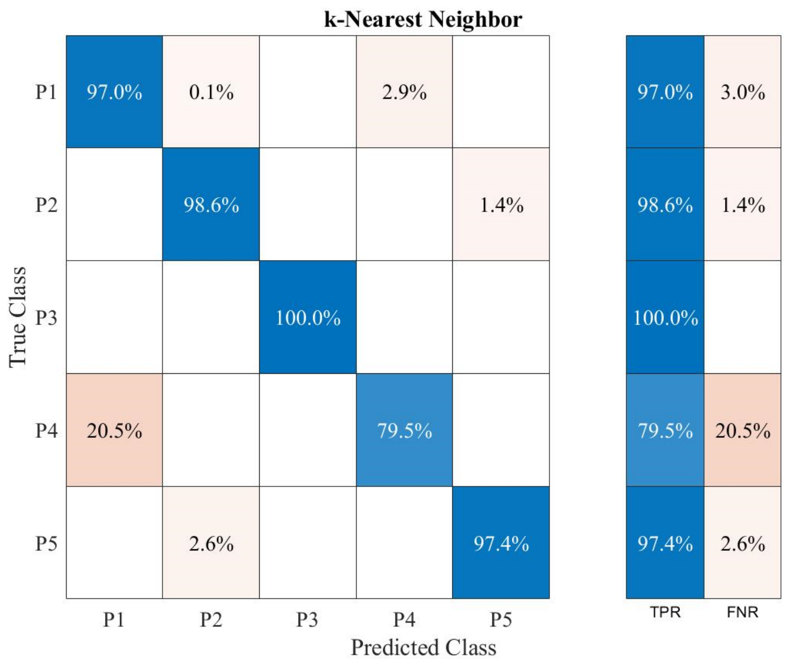


Figure 3.46: The confusion matrix of the k-nearest neighbor model for Case 4.

Similar results with Case 2 and 3 are observed for Case 4 too regarding k-nearest neighbor model predictions. The highest accuracy, 95.7%, is achieved by selecting the k number as 50. As in Case 2 and 3, this state results in the overfitting at its lowest in the model variation where the k number is selected as 50. The lowest accuracy, 94.5%, is obtained for the selection of k number as 1. The reason is the overfitting of the model on the training set.

In Figure 3.46, the confusion matrix of the k-nearest neighbor model for Case 4 is given. In similar, the highest FNR value is observed between the P1 and P4. 20.5% of the data classified as P4 is predicted as P1.

Table 3.11: The accuracies of Case 5 according to model features

Models	Features	Accuracy (%)
Decision Tree (Split criterion=Gini's Index)	Maximum number of splits=100	99.7
	Maximum number of splits=85	99.7
	Maximum number of splits=50	99.7
	Maximum number of splits=25	99.7
Kernel Naïve Bayes (Kernel type=Triangle)	Support=Positive	99.8
	Support=Unbounded	99.0
k-nearest neighbors (Distance metric=Euclidean)	Number of neighbors=1	99.9
	Number of neighbors=10	99.9
	Number of neighbors=15	99.9
	Number of neighbors=50	99.9
	Number of neighbors=100	99.2
	Number of neighbors=150	90.9

For Case 5, the responses are selected as Case 1 and 2. In difference, the pressure drop values of CH and DHW channels are selected as predictors. The training data distribution is shown in Figure 3.47. The data are clearly distinguishable from each other. Therefore, the accuracies are too high compared to the other cases. The models tend to overfit the training data.

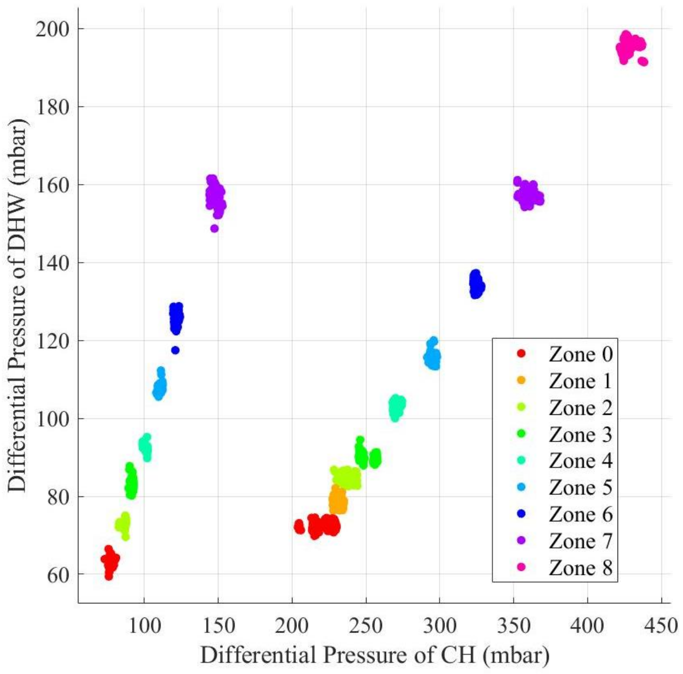


Figure 3.47: The training data distribution shown in the pressure difference of CH and DHW for Case 5.

In the decision tree model for Case 5, the accuracies of the model variations are obtained as same. The confusion matrix of the decision tree model with the selection of a maximum number of 25 is given in Figure 3.48.

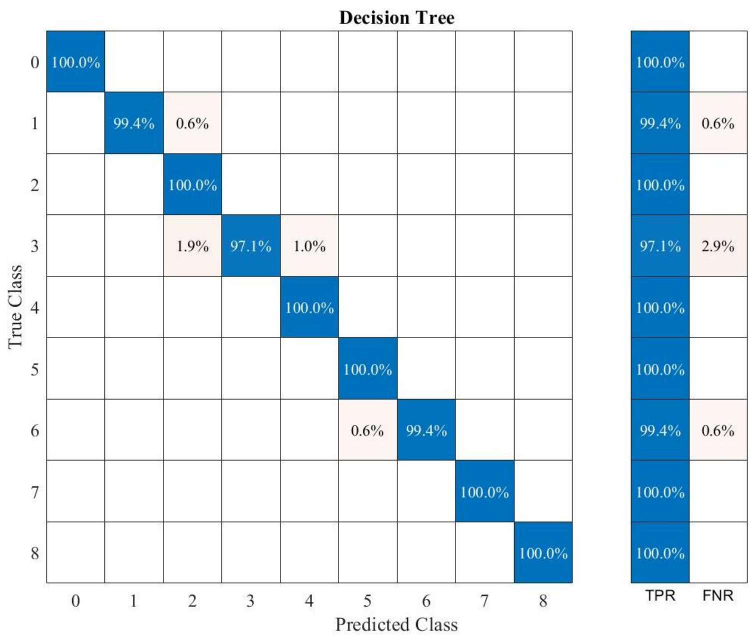


Figure 3.48: The confusion matrix of the decision tree model for Case 5.

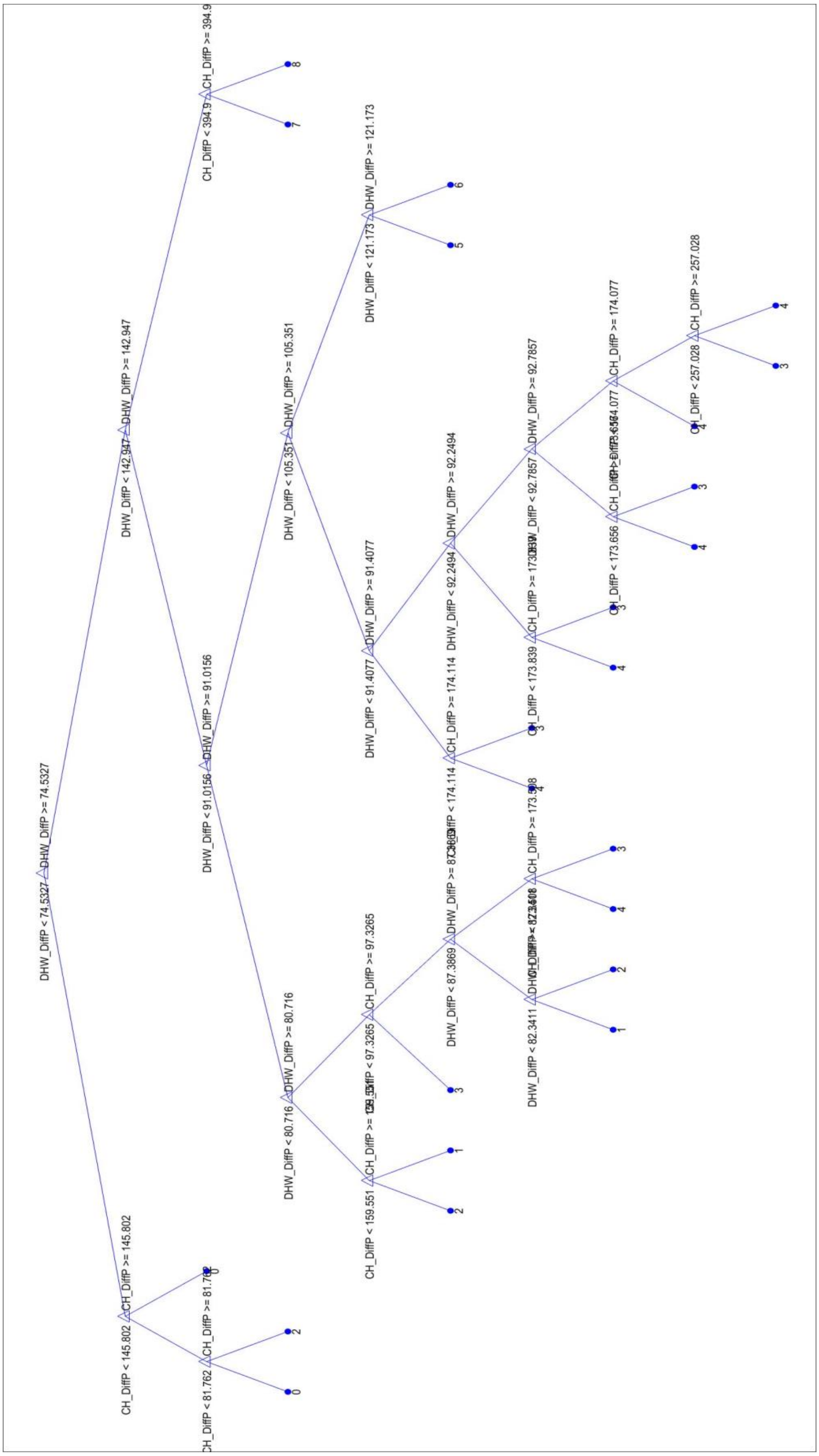


Figure 3.49: The classification tree designation of the decision tree model for Case 5.

The classification tree designation for Case 5 is given in Figure 3.49. The differential pressure values of CH and DHW channels are easy to distinguish from each other, as shown in Figure 3.50. Therefore, the classification of the testing data based on the trained model is achieved with high accuracy. This is the demonstration of the high similarity between the training data and testing data.

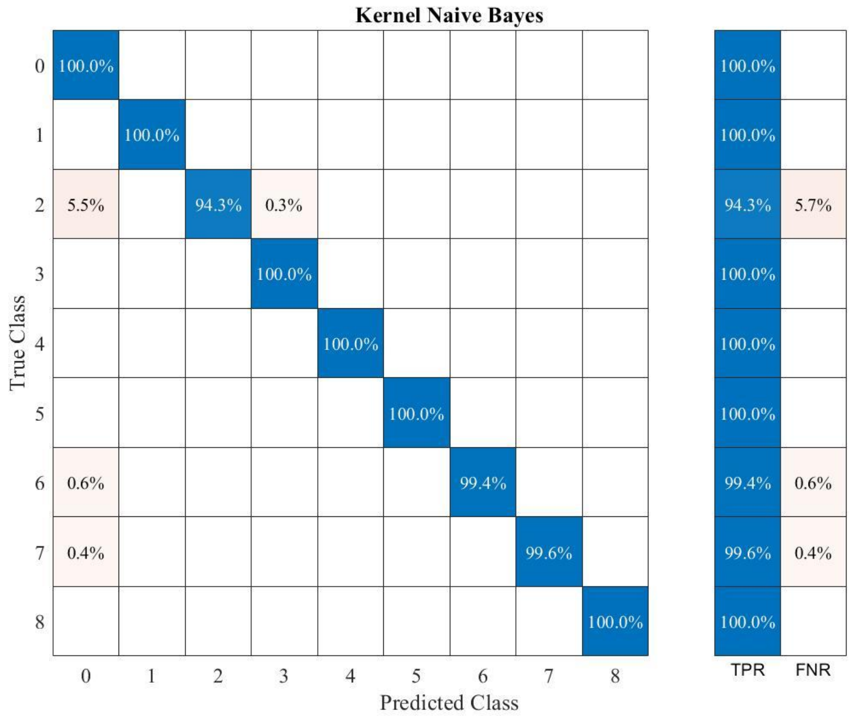


Figure 3.50: The confusion matrix of Kernel Naïve Bayes for Case 5.

The Kernel Naïve Bayes model with the selection of unbounded support function gives 99.0% of accuracy, which is slightly less than the model variation with the positive support function. The mentioned reason for the high accuracy is valid for this model too. The confusion matrix of the Kernel Naïve Bayes model is given in Figure 3.50. The highest FNR value is obtained between Zone 2 and Zone 0. 94.3% of the data classified as Zone 2 is predicted correctly. 5.5% of it is predicted as Zone 0 and 0.3% of it is predicted as Zone 3, with a total FNR of 5.7%.

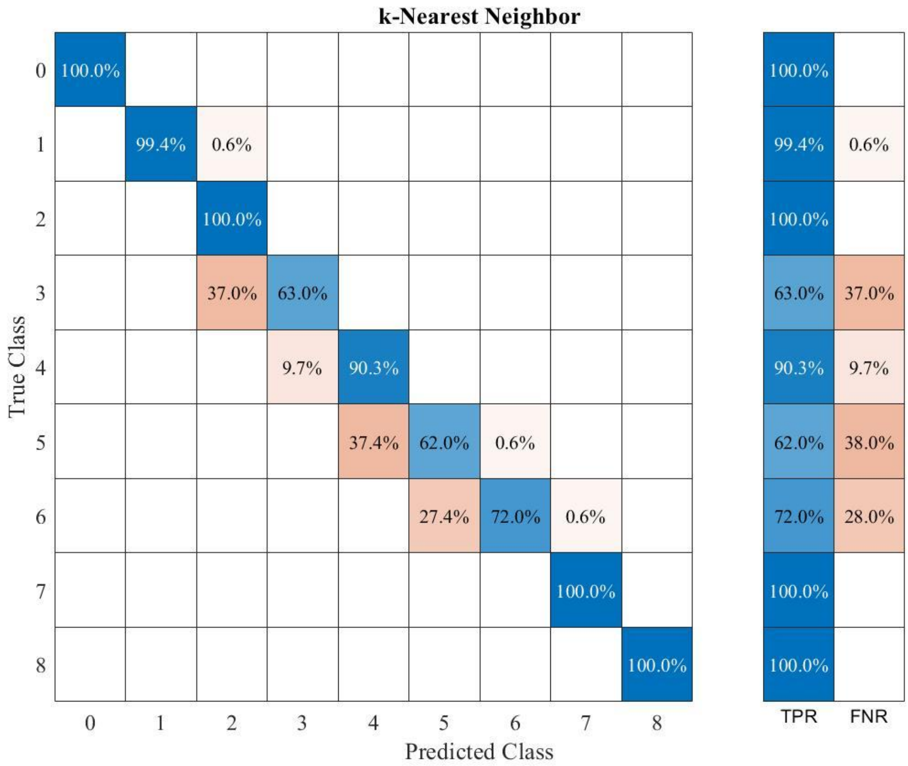


Figure 3.51: The confusion matrix of the k-Nearest Neighbor model for Case 5.

The k-nearest neighbor model gives the same and nearly perfect prediction accuracy for the variants of the model that has been trialed before for Case 1 to 4. The k-nearest neighbor model tends to overfit the training data set. The nearly perfect prediction accuracy, 99.9%, is obtained with these predictor selections and the applied cross-validation method. The two variants of this model are selected with k number 100 and 150. The lowest accuracy is seen for the model with the selection of 150 for k number, in Table 3.11.

The confusion matrix of the k-nearest neighbor model with the selection of 150 for k number is given in Figure 3.51. The highest FNR value is obtained between Zone 5 and Zone 4. 37.4% of the data classified as Zone 5 is predicted as Zone 4. It is followed by the false prediction of 37.0% of the data classified as Zone 3, as Zone 2. Even though the FNR value is not too low compared to other cases, false predictions are encountered for close classes.

Table 3.12: The accuracies of Case 6 according to model features

Models	Features	Accuracy (%)
Decision Tree (Split criterion=Gini's Index)	Maximum number of splits=100	98.7
	Maximum number of splits=85	98.7
	Maximum number of splits=50	98.7
	Maximum number of splits=25	94.7
Kernel Naïve Bayes (Kernel type=Triangle)	Support=Positive	89.1
	Support=Unbounded	89.0
k-nearest neighbors (Distance metric=Euclidean)	Number of neighbors=1	98.1
	Number of neighbors=10	91.1
	Number of neighbors=15	86.9
	Number of neighbors=50	70.8

For the last Case 6, the responses are kept the same as Case 5, the only difference is the predictors. The CH inlet temperature, DHW outlet temperature and DHW flow rate features have already been measured by the combi-boiler control unit during the in-real-life operation, without the need for any additional sensor or equipment. Therefore, they are used as the predictors for Case 6.

As in Case 1 and 5, the obtained accuracies of the decision tree model with the selection of maximum number of splits 100, 85 and 50 are the same as each other within the cases. However, the model with the selection of the maximum number of splits as 25 is obtained as differently lower from the others.

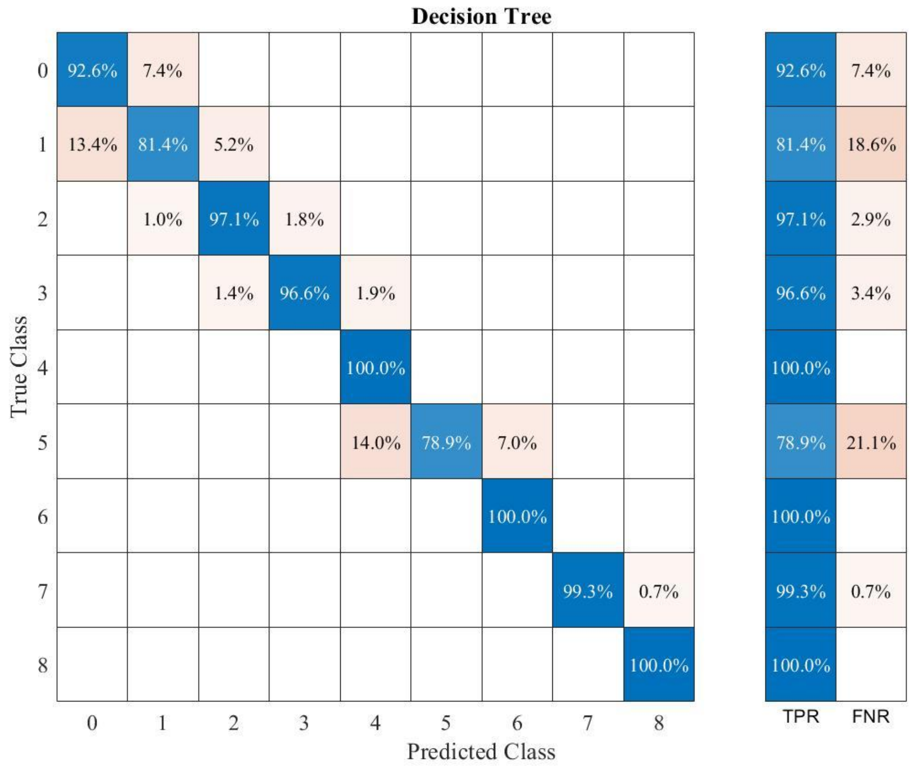


Figure 3.52: The confusion matrix of the decision tree model for Case 6.

The confusion matrix of the decision tree model with the selection of the maximum number of splits as 25 is given in Figure 3.52. Similar to Case 5, the FNR values are obtained between the close classes. The highest FNR value is obtained between Zone 5 and 4. 14.0% of the data classified as Zone 5 is predicted as Zone 4, 7.0% of the data classified as Zone 5 is predicted as Zone 6.

The classification tree designation of the decision tree model is given in Figure 3.53. In Figure 3.53, the classification node is started with the decision rule as classifying Zone 8 first. This shows Zone 8 has more clear distinctions than the others.

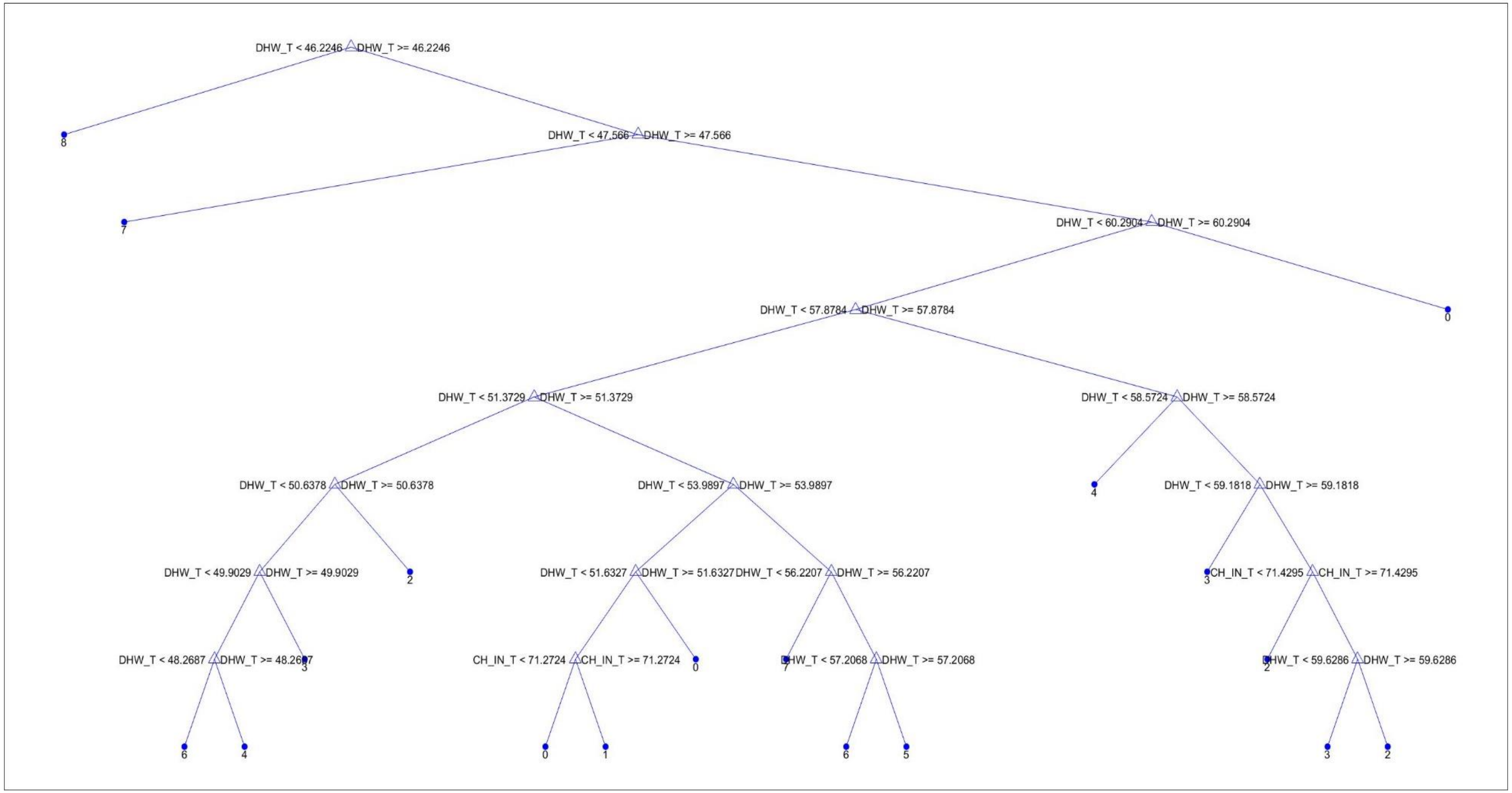


Figure 3.53: The classification tree designation of the decision tree model for Case 6.

The Kernel Naïve Bayes models with two selections of support function type, give approximately the same results, as in Case 1. The obtained accuracy for the unbounded support function is 89.0%, while it is 89.1% for the positive support function. The confusion matrix of the Kernel Naïve Bayes model that is used with unbounded support function is given in Figure 3.54. The highest FNR value is obtained for Zone 0 and Zone 1. 25.0% of the data classified as Zone 0 is predicted as Zone 1. If customer comfort is considered, between Zone 0 and Zone 1 there is a mild comfort difference. Therefore, the prediction accuracy of classes close to the worst zone, Zone 8, is preferred.

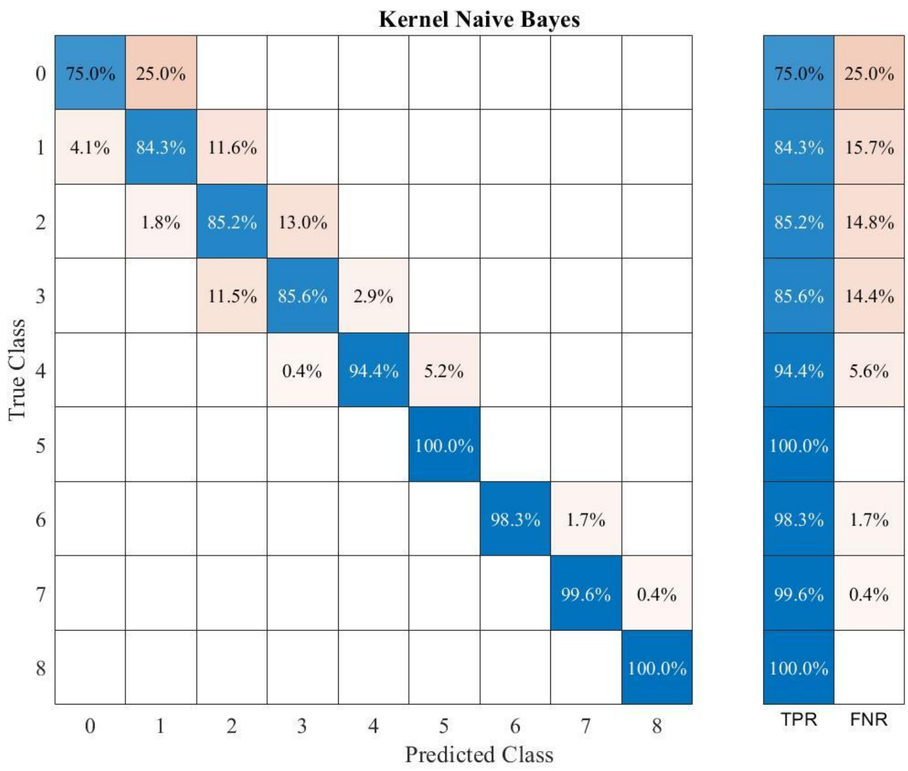


Figure 3.54: The confusion matrix of Kernel Naïve Bayes model for Case 6.

In contrast to the other cases, the k-nearest neighbor model gives lower accuracy for higher k number. This means the best result is only achieved when the model is too fitted to the training data. The lowest accuracy of the k-nearest neighbor model is obtained as 70.8, for k number is selected as 50. The highest accuracy is achieved as 98.1, with the selection of k number as 1.

The confusion matrix of the k-nearest neighbor model for the selected k number as 50 is given in Figure 3.55. The highest FNR value is obtained between Zone 1 and Zone 0, similar to the Kernel Naïve Bayes model. The 39.0% of the data that is classified as Zone 1 is predicted as Zone 0. In contrast of the other models, there are classes that the true prediction percentages are not higher than the false ones.

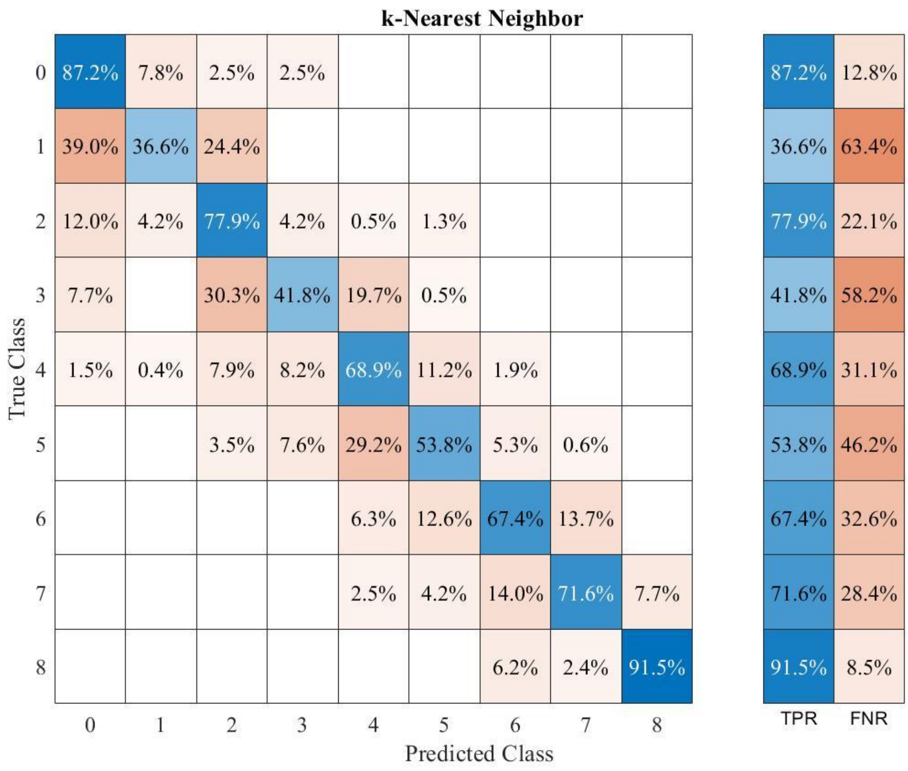


Figure 3.55: The confusion matrix of the k-Nearest Neighbor model for Case 6.

Chapter 4

Conclusion

Fouling is an accumulation of undesired particles on the heat surfaces that causes a lack of heat transfer. The fouling of plate heat exchangers, which is used in combi-boilers is investigated in this thesis. The main aim of the thesis is to investigate the machine learning algorithms to classify and predict the fouling status of PHE used in combi-boilers, to generate the background of the predictive maintenance that is willing to apply to combi-boilers control unit, besides investigating the fouling effect on PHEs in terms of heat transfer and energy consumption by using a 1-D model.

The artificially generated method of experiments is used to obtain the data that is required for the algorithm training. The data obtained from experiments are representing the fouling behavior of the PHEs that have 30 and 32 plates. These obtained data show the fouling effects on PHE can be observed by the used method. The effect of fouling on PHE performance is assumed as similar to the performance loss that would be occurred if the PHE that is already used in the combi-boiler, i.e., is already designed for the combi-boiler, is replaced with a PHE that has fewer plate numbers. The plate numbers, and the size of the PHE, is designed according to the combi-boilers required power output. With this assumption, the experiment results show that the expected trends of output temperatures and pressure drop values of both channels are seen.

The overall heat transfer coefficient and fouling resistance coefficient are calculated as the performance values of the tested PHEs. As expected, the overall heat transfer coefficients are resulted in decreasing while the fouling resistance coefficient is increasing. The values are compared to the reference, which is a study that the particulate and composite fouling of PHEs is investigated by adding the particulates as an accelerated test. The obtained trends and values demonstrate similar results with the reference study in the literature.

A 1-D numerical model is structured by using Runge Kutta 4^{th} order ordinary differential equation solving method. The differential equations are generated according to the thermal resistance method based on energy balance. Temperature distribution of 32 and 30 plate PHEs by time and position was examined in a model created in MATLAB. Three validation studies are generated to verify the model and experiment results. The created model has almost similar results to the referenced study [33]. The model results are also compared with the healthy and faulty experimental results. The errors for healthy values is less than 2%. The temperature difference at maximum for faulty values are less than 2°C. Therefore, the model is considered as correct.

Due to fouling layer occurrence on the plate surfaces, the combi-boiler appliances need more natural gas to heat the CH line to reach the required setpoint of DHW outlet temperature. This additional required power to reach the setpoint of DHW defined by the customer is calculated at maximum fouling by using the model results. The results show that combi-boiler appliances need to supply approximately 16 and 7 kW additional heat output to reach the required setpoint of DHW in case of maximum fouling.

The obtained data from experiments are used to be implied to the Classification Learner Application by MATLAB. Different cases are created to investigate the model performances regarding the predictor and response selection. The Naïve Bayes, k-nearest neighbors, and decision tree models are used. The models are trained according to the experiment data grouped by classes regarding customer comfort and test conditions. The training data and testing data splitting is generated by using the k-fold cross-validation method to avoid overfitting.

The results show that each algorithm gives a considerable performance in each case. The k-nearest neighbors model gives higher prediction accuracies than the other models except for Case 6. However, the k-nearest neighbors algorithm tends to overfit the training data, while the selected number of neighbors is decreasing. Therefore, the best result of k-nearest neighbor is encountered for the Case 2,3 and 4 where the predictor is selected as the overall heat transfer coefficient, due to the increasing accuracy regarding with increasing number of neighbors. It results that the k-nearest neighbors model would be the best among the other models for predicting the classes according to the overall heat transfer coefficient values. The decision tree model results show that the model is independent of its maximum number of splits selection. The model achieves approximately the same accuracies even the maximum number of splits value is changed. The results show the decision tree model gives better performance in classifying than the Naïve Bayes model according to the accuracy results. In further studies, the selected models would be tested with the real data obtained over time to see the results of the integrated model on combi-boilers.

References

[1] Hewitt GF, Shires GL, Bott TR. Process Heat Transfer. Boca Raton, FL: CRC Press; 1994.

[2] Song KS, Lim J, Yun S, Kim D, Kim Y. Composite fouling characteristics of CaCO3 and CaSO4 in plate heat exchangers at various operating and geometric conditions. Int J Heat Mass Transf [Internet]. 2019;136:555–62. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2019.03.032

[3] Lee E, Jeon J, Kang H, Kim Y. Thermal resistance in corrugated plate heat exchangers under crystallization fouling of calcium sulfate (CaSO4). Int J Heat Mass Transf [Internet]. 2014;78:908–16. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2014.07.069

[4] Pääkkönen TM, Riihimäki M, Simonson CJ, Muurinen E, Keiski RL. Crystallization fouling of CaCO3 – Analysis of experimental thermal resistance and its uncertainty. Int J Heat Mass Transf [Internet]. 2012;55(23–24):6927–37. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2012.07.006

[5] Pääkkönen TM, Riihimäki M, Simonson CJ, Muurinen E, Keiski RL. Modeling CaCO3 crystallization fouling on a heat exchanger surface – Definition of fouling layer properties and model parameters. Int J Heat Mass Transf [Internet]. 2015;83:84–98. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2014.11.073

[6] Zhang G-M, Li G-Q, Li W, Zhang Z, Leng X-L, Tian M-C. Particulate fouling and composite fouling assessment in corrugated plate heat exchangers. Int J Heat Mass Transf [Internet]. 2013;60:263–73. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2013.01.040

[7] Pekgüzelsu M. Bir Kombide Kullanılan Plakalı Isı Değiştiricisinde Birikinti Oluşumunun Deneysel Olarak İncelenmesi. [İzmir]: Ege Üniversitesi; 2015.

[8] Vargel C. Corrosion of Aluminium. London, England: Elsevier Science; 2020.

[9] Davis JR, editor. Corrosion of aluminum and Aluminum Alloys. ASM International; 1999.

[10] Gustafsson S. Corrosion properties of aluminium alloys and surface treated alloys in tap water. [Uppsala]: Uppsala Universitet; 2011.

[11] Szklarska-Smialowska Z. Pitting corrosion of aluminum. Corros Sci [Internet]. 1999;41(9):1743–67. Available from: http://dx.doi.org/10.1016/s0010-938x(99)00012-8

[12] Nguyen H-P, Fauriat W, Zio E, Liu J. A data-driven approach for predicting the remaining useful life of steam generators. In: 2018 3rd International Conference on System Reliability and Safety (ICSRS). IEEE; 2018.

[13] Ardsomang T, Wesley Hines J, R. Upadhyaya B. Heat exchanger fouling and estimation of Remaining Useful Life. Proc Annu Conf Progn Health Manag Soc [Internet]. 2021;5(1). Available from: http://dx.doi.org/10.36001/phmconf.2013.v5i1.2773

[14] Sun L, Zhang Y, Saqi R. Research on the fouling prediction of heat exchanger based on Support Vector Machine optimized by Particle Swarm Optimization algorithm. In: 2009 International Conference on Mechatronics and Automation. IEEE; 2009.

[15] Sun L, Zhang Y, Zheng X, Yang S, Qin Y. Research on the fouling prediction of heat exchanger based on support vector machine. In: 2008 International Conference on Intelligent Computation Technology and Automation (ICICTA). IEEE; 2008.

[16] Riverol C, Napolitano V. Estimation of fouling in a plate heat exchanger through the application of neural networks. J Chem Technol Biotechnol [Internet]. 2005;80(5):594–600. Available from: http://dx.doi.org/10.1002/jctb.1198

[17] Mohanty DK, Singru PM. Fouling analysis of a shell and tube heat exchanger using local linear wavelet neural network. Int J Heat Mass Transf [Internet]. 2014;77:946–55. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2014.06.007

[18] Wen X, Miao Q, Wang J, Ju Z. A multi-resolution wavelet neural network approach for fouling resistance forecasting of a plate heat exchanger. Appl Soft Comput [Internet]. 2017;57:177–96. Available from: http://dx.doi.org/10.1016/j.asoc.2017.03.043

[19] Sundar S, Rajagopal MC, Zhao H, Kuntumalla G, Meng Y, Chang HC, et al. Fouling modeling and prediction approach for heat exchangers using deep learning. Int J Heat Mass Transf [Internet]. 2020;159(120112):120112. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2020.120112

[20] Dheenamma M, Soman DP, Muthamizhi K, Kalaichelvi P. In pursuit of the best artificial neural network configuration for the prediction of output parameters of corrugated plate heat exchanger. Fuel (Lond) [Internet]. 2019;239:461–70. Available from: http://dx.doi.org/10.1016/j.fuel.2018.11.034

[21] Palsson OP, Jonsson GR, Desmet B, Lalot S. Comparison of neural networks and Kalman filters performances for fouling detection in a heat exchanger. International Journal of Heat Exchangers. 2007;VII(1524–5608):151–68.

[22] Wang W, Christer AH. Towards a general condition based maintenance model for a stochastic dynamic system. J Oper Res Soc [Internet]. 2000;51(2):145. Available from: http://dx.doi.org/10.2307/254254

[23] Sun J, Wang F, Ning S. Aircraft air conditioning system health state estimation and prediction for predictive maintenance. Chin j aeronaut [Internet]. 2020;33(3):947–55. Available from: http://dx.doi.org/10.1016/j.cja.2019.03.039

[24] Shohet R, Kandil MS, McArthur JJ. Machine learning algorithms for classification of boiler faults using a simulated dataset. IOP Conf Ser Mater Sci Eng [Internet]. 2019;609(6):062007. Available from: http://dx.doi.org/10.1088/1757-899x/609/6/062007

[25] Aggarwal CC, editor. Data Classification: Algorithms and Applications. London, England: CRC Press; 2020.

[26] Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms. Cambridge, England: Cambridge University Press; 2014.

[27] Awad M, Khanna R. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. 1st ed. Berlin, Germany: APress; 2014.

[28] Mohri M, Rostamizadeh A, Talwalkar A. Foundations of machine learning. 2nd ed. London, England: MIT Press; 2018.

[29] Bradley P. Machine learning - A comprehensive, step-by-step guide to learning and applying advanced concepts and techniques in machine learning. Draft2digital; 2019.

[30] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction, second edition. 2nd ed. New York, NY: Springer; 2017.

[31] Blum A, Hopcroft J, Kannan R. Foundations of Data Science. Cambridge, England: Cambridge University Press; 2020.

[32] Tosun İ. Modeling in Transport Phenomena. Elsevier; 2007.

[33] Bobič M, Gjerek B, Golobič I, Bajsić I. Dynamic behaviour of a plate heat exchanger: Influence of temperature disturbances and flow configurations. Int J Heat Mass Transf [Internet]. 2020;163(120439):120439. Available from: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2020.120439

Curriculum Vitae

Name Surname : Ceren Vatansever

Education:

2015–2019 İzmir Kâtip Çelebi University, Dept. of Mechanical Eng., Bachelor's Degree

2019–2022 İzmir Kâtip Çelebi University, Graduate School of Natural and Applied Sciences, Master of Science Degree

Work Experience:

2019 – 2021 CFM Soğutma ve Otomasyon A.Ş, R&D Engineer

2021-present Bosch Termoteknoloji Isıtma ve Klima Sanayi Ticaret A.Ş., Hydraulic Development Engineer

Publications:

1. Kuzucanlı SA, Vatansever C, Yaşar AE, Haktan Karadeniz Z. Assessment of fouling in plate heat exchangers using classification machine learning algorithms. CLIMA [Internet]. 2022 May 13 [cited 2022 Jul. 3];. Available from: https://proceedings.open.tudelft.nl/clima2022/article/view/127