Applied Thermal Engineering 210 (2022) 118305

Contents lists available at ScienceDirect

Applied Thermal Engineering

journal homepage: www.elsevier.com/locate/apthermeng

Consolidated modeling and prediction of heat transfer coefficients for saturated flow boiling in mini/micro-channels using machine learning methods

Ari Bard [a], Yue Qiu [a], Chirag R. Kharangate [a,*], Roger French [b]

[a] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
[b] Department of Material Science, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA

ARTICLE INFO

Keywords:
Machine learning
Neural networks
Support vector machines
Flow boiling
Heat transfer

ABSTRACT

Flow boiling has become a reliable mode of compensating with larger power densities and greater functions of devices because it is able to utilize both the latent and sensible heat contained within a specified coolant. There are currently very few available tools proven reliable when predicting heat transfer coefficients during flow boiling in mini/micro-channels. The most popular methods rely on semi-empirical correlations derived from experimental data. These correlations can only be applied to a very narrow subset of testing conditions. This study uses a number of data science methods and techniques to accurately predict the heat transfer coefficient during flow boiling in mini/micro-channels on a database consisting of 16,953 observations collected across 50 experiments using 12 working fluids. Exploratory data science is used to obtain confidence in the data and investigate relationships between feature variables before employing machine learning algorithms. Missing data is imputed using random forest nonparametric imputation. A variety of feature analysis techniques are employed to combine and select different optimal feature variables as input values such as principal component analysis to reduce the overall dimensionality of the dataset and the Boruta package, recursive feature elimination, Least Absolute Shrinkage and Selection Operator (LASSO) regression, and stepwise selection to reduce the number of original variables used when modeling while preserving as much information as possible. A variety of models including linear modeling, generalized additive modeling, random forests, support vector machines, and neural networks are used to predict the heat transfer coefficient and compare the results with existing universal correlations. The support vector machine model performed best, with a Mean Absolute Percentage Error (MAPE) of 11.3%. The heat flux, vapor-only Froude number, and quality proved to be especially significant contributing variables across 90% of over 110 different models. Machine learning proved to be an extremely useful tool when predicting the heat transfer coefficient across a variety of different fluids but did struggle to predict extremely high outlier data where water was the working fluid.

1. Introduction

1.1. Mini and micro-channels for thermal management

The rapid advancement of high performance electronic systems over the past several decades has come with an increasing demand for better thermal management systems. Traditional thermal management systems use single-phase liquid or air flows to meet the cooling needs. Flow boiling in mini/micro-channels has become a reliable mode of adapting to larger power densities because it is able to utilize both the latent and sensible heat contained within a specified coolant. Mini/Micro-channels have seen popular use when dealing with high density systems because of their increased capacity for temperature uniformity, their high power density as a result of their high surface area-to-volume ratio, and their lower liquid requirement than other traditional modes of convective heat transfer due to the normally required much lower mass flow rates [1]. In addition, the smaller surface temperature variations, can also reduce thermo-mechanical stresses on the objects being cooled and improve durability. With advances over the past few decades, mini/micro-channels have also become easy to manufacture in both single and multi-channel configurations of various cross-sectional geometries [2,3]. Their primary drawbacks include high pressure drops and a propensity for flow instability and two-phase choking. Another drawback is

* Corresponding author.
E-mail address: chirag.kharangate@case.edu (C.R. Kharangate).

https://doi.org/10.1016/j.applthermaleng.2022.118305
Received 8 November 2021; Received in revised form 3 February 2022; Accepted 2 March 2022
Available online 15 March 2022

Nomenclature

Bd	Bond number, $Bd = g(\rho_f - \rho_g)D_h^2/\sigma$
Bo	Boiling Number, $Bo = q_H''/Gh_{fg}$
Co	Convection number, $Co = [(1-x)/x]^{0.8}(\rho_g/\rho_f)^{0.5}$
c_p	Specific heat at constant pressure, [kJ/(kgK)]
c_v	Specific heat at constant volume, [kJ/(kgK)]
D_h	Hydraulic diameter of flow channel, [mm]
Fr_f	Saturated liquid Froude number, $Fr_f = [G(1-x)]^2/(\rho_f^2 g D_h)$
Fr_g	Saturated vapor Froude number, $Fr_g = (Gx)^2/(\rho_g^2 g D_h)$
Fr_{fo}	Liquid-only Froude number, $Fr_{fo} = G^2/(\rho_f^2 g D_h)$
Fr_{go}	Vapor-only Froude number, $Fr_{go} = G^2/(\rho_g^2 g D_h)$
G	Mass velocity, [kg/m^{2}s]
g	Gravity acceleration, [m/s^2]
h	Heat transfer coefficient, [kW/(m^{2}K)]
h_{fg}	Latent heat of vaporization, [kJ/(kgK)]
k	Thermal conductivity, [W/(mK)]
L	Channel length, [mm]
M	Molecular mass, [kg/mol]
MAE	Mean absolute error
$MAPE$	Mean absolute percentage error
MSE	Mean square error
n	Number of input parameters
P	Saturation pressure, [Pa]
P_c	Critical pressure, [Pa]
Pe_f	Saturated liquid Peclet number, $Pe_f = Re_f Pr_f$
Pe_g	Saturated vapor Peclet number, $Pe_g = Re_g Pr_g$
P_F	Wetted perimeter of channel, [mm]
P_H	Heated perimeter of channel, [mm]
P_R	Reduced pressure, $P_R = P_c/P$
Pr_f	Saturated liquid Prandtl number, $Pr_f = \mu_f c_{pf}/k_f$
Pr_g	Saturated vapor Prandtl number, $Pr_g = \mu_g c_{pg}/k_g$
q''	Heat flux, [kJ/m^2]
q_H''	Heat flux based on heated perimeter of channel, [kJ/m^2]
R	Relative roughness, $R = e/D_h$
R^2	Coefficient of determination
Re_f	Saturated liquid Reynolds number, $Re_f = G(1\text{-}x)D_h/\mu_f$
Re_g	Saturated vapor Reynolds number, $Re_g = GxD_h/\mu_g$
Re_{fo}	Liquid-only Reynolds number, $Re_{fo} = GD_h/\mu_f$
Re_{go}	Vapor-only Reynolds number, $Re_{go} = GD_h/\mu_g$
$RMSE$	Root Mean Square Error
Su_f	Saturated liquid Suratman number,
Su_g	Saturated vapor Suratman number,
T	Saturation temperature, [K]
We_f	saturated liquid Weber number, $We_f = [G(1-x)]^2 D_h/(\rho_f \sigma)$
Weg	saturated vapor Weber number, $We_g = (Gx)^2 D_h/(\rho_g \sigma)$
$Wefo$	liquid-only Weber number, $We_{fo} = G^2 D_h/(\rho_f \sigma)$
$Wego$	vapor-only Weber number, $We_{go} = G^2 D_h/(\rho_g \sigma)$
X	Mole Fraction
x	Quality, generic variable
X_{tt}	Lockhart-Martinelli parameter, $X_{tt} = \left(\frac{\mu_f}{\mu_g}\right)^{0.1}\left(\frac{1-x}{x}\right)^{0.9}\left(\frac{\rho_g}{\rho_f}\right)^{0.5}$

Greek Symbol

α	Aspect ratio
β	Weighted coefficient
λ	Weighted coefficient error term
μ	Dynamic viscosity
Φ	Principal component weighting factor
ρ	Density
σ	Surface tension

Subscripts

cb	Convective boiling
di	After diffusion
f	Saturated liquid, fluid
FB	Flow boiling
fo	Liquid only
g	Saturated vapor
go	Vapor only
nb	Nucleate boiling
obs	Observed
$pred$	Predicted
tp	Two phase
w	Channel wall
Tt	turbulent liquid-turbulent vapor

Acronyms

C	Circular
R	Rectangular
H	Horizontal
VU	Vertical Upflow
$C+L$	Copper + Lexan
$S+L$	Silicon + Lexan
SS	Stainless Steel
CB	Convective Boiling
NB	Nucleate Boiling

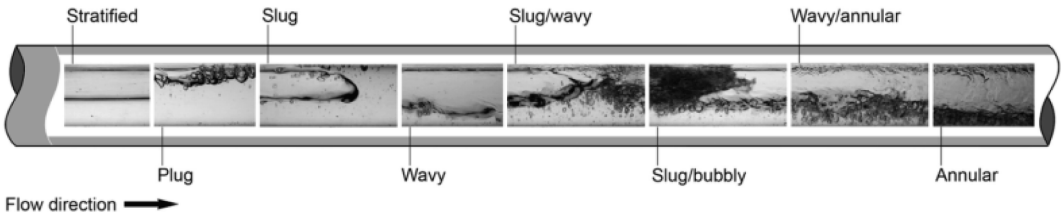

Fig. 1. Different strata of two phase flow. Adapted from: Ghajar et al. [5].

that flow boiling in mini/micro-channels is also a very different phenomenon than in macro-channels where many experiments [4,5] and modelling [6–8] has been conducted [9]. Current delineations in channel size are often arbitrary and do not account for a difference in flow behavior. Additionally, established experimental research to predict fluid characteristics on a macro-channel level does so under turbulent flow conditions, but most conditions on a mini/micro-scale level have liquid Reynolds numbers below 2300 and are thus laminar [3,10]. As a result, very little macro-channel flow data can be considered when looking at flow boiling in mini/micro-channels.

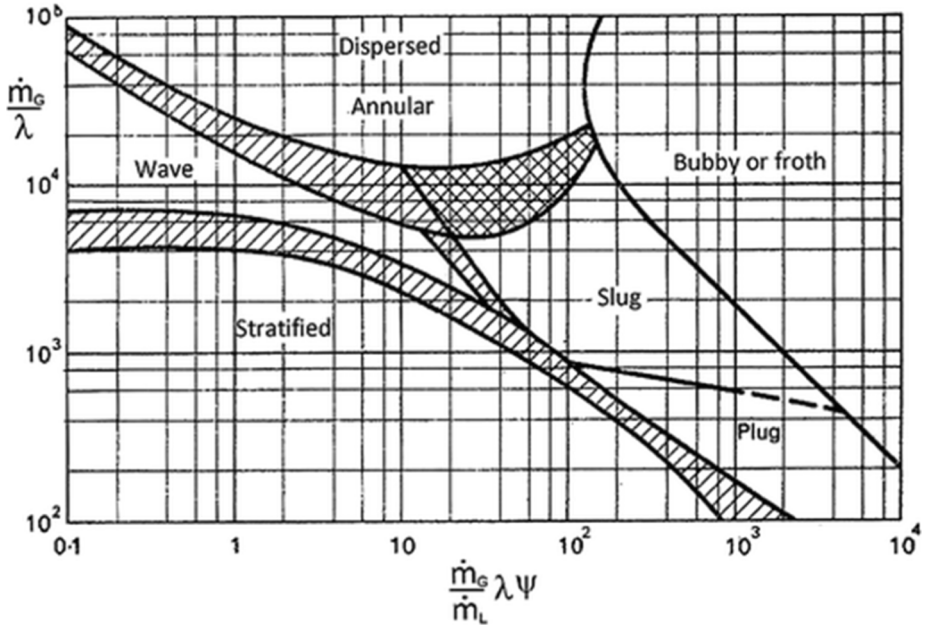

Fig. 2. Flow-Pattern Map. Adapted from: Yadigaroglu et al. [6].

1.2. Flow boiling regimes and heat transfer

Flow boiling behavior is categorized into the different types as shown in Fig. 1 based on the presence of bubble formation, coalescence, slug or plug formation, or dryout conditions and has often been determined with the help of flow-pattern maps such as the one depicted in Fig. 2 [10,11]. These maps chart the flow behavior on a two-dimensional plane as a function of different transition criteria but are often limited to adiabatic conditions. Local flow boiling behavior is often necessary for design parameters, and many applications use diabatic flow boiling. There are many different possible flow regimes that can occur. During boiling, the fluid often starts as a pure liquid. As trace amounts of vapor begin to appear, small bubbles start to form. The increase in the flow rate of that vapor leads to plug flow, and when long, elliptical bubbles form in the center and along the length of the channel, slug flow occurs [12]. When the shear forces exerted by the vapor center propel the liquid to a very thin area along the walls of the channel, annular flow occurs. If both gas and liquid flow rates increase, the flow will become an unstable churn flow regime pattern. At very high gas flow rates, the liquid film on the walls of the channel could be aerosolized into droplets by the fast-moving gas in a state known as mist flow [13]. Additionally, Cubaud and Ho [14] have identified a wedge flow behavior that exists as an intermediary state between bubbly and slug flow where the liquid film on the walls of the channel has dry patches which can significantly affect pressure drops and thin film evaporation. The pattern behaviors are often determined subjectively and can follow numerous naming conventions or employ a multitude of subcategories [10]. A way to standardize flow boiling behavior would likely prove valuable in future research. These different flow behaviors largely describe different ranges of vapor quality, and numerous experiments have shown that key heat transfer trends in microchannels fluctuate dramatically at different vapor qualities. For example, at lower vapor qualities, the heat transfer coefficient has a slightly negative or no relationship to quality, but at higher qualities, the heat transfer coefficient has a strong negative relationship to quality [15]. Both the Bertsch [101] and Kim & Mudawar [102] semi-empirical correlations used as baselines for comparison in this study contain vapor quality terms when calculating the heat transfer coefficient. Flow boiling instabilities can cause large amplitude oscillations in mass flow rate, pressure, temperature, and other fluid parameters and can significantly influence the local heat transfer coefficient. Channel surface parameters, surface wettability, thermal conductivity, or size of the inlet/outlet manifold can also act as confounding properties that cause flow instability by causing unanticipated rapid nucleation or bubble growth [16]. The channel inlets and outlets can be altered in some cases to reduce the amount of flow instabilities, but it is unclear as to whether these instabilities always lead to lower rates of heat transfer. Therefore, the variations in flow behavior and corresponding heat transfer makes it difficult to predict the heat transfer coefficients in mini/micro-channel.

1.3. Predicting flow boiling data in mini/micro-channels

There are currently very few available tools or methods proven reliable when predicting heat transfer coefficients during flow boiling in mini/micro-channels. The most popular methods rely on empirical and semi-empirical correlations [17,18] derived from experimental data. These correlations can mostly only be applied to a very narrow subset of testing conditions for a very limited variety of fluids, channel sizes, and channel geometries [14]. Currently, the results can produce different and contradicting trends with high errors if tested outside the development range based on the experimental setup and procedure used by different researchers. These errors can be the result of experimental and measurement inaccuracies of input parameters, known as parameter uncertainty, or errors in the prediction itself due to approximations made when creating the model known as form uncertainty. Another predicting method to predict the heat transfer coefficients is the use of theoretical models based on the physical behavior of the flow. For example, the annular flow condensation model developed by Kim and Mudawar [19] can predict steady-state variations of fluid flow and thermal behavior including heat transfer coefficient. However, its applicability is narrow and limited to annular flow regime. With the recent developments in computing techniques, computational fluid dynamics (CFD) simulations [20–23] has become very promising comparing with traditional techniques. The advantages of CFD approaches are capability of predicting transient flow and heat transfer behavior with detailed information of void fraction, phase velocities and temperatures. However, the low accuracy due to model development and the high computational cost have limited the utilization of CFD methods [24].

Machine learning can help identify and revise elements causing

Table 1
Displays flow boiling heat transfer data organized by experiment along with relevant experiment setup characteristics [21–70].

Author(s)	Channel geometry[a]	Channel material	D_h [mm]	Relative roughness R, e/D_h	Fluid(s)	G [kg/m^{2}s]	Data points
Wambsganss *et al.* (1993)	C single, H	Stainless steel	2.92	Smooth	R113	50–300	92
Tran (1998)	C single, H	Brass	2.46	Smooth	R134a	33–502	302
Wang *et al.* (1998)	C single, H	Copper	6.5	Smooth	R22	100–400	63
Yan and Lin (1998)	C multi, H	Copper	2.0	–	R134a	50–200	137
Bao *et al.* (2000)	C single, H	Copper	1.95	Smooth	R11, R123	167–560	164
Qu and Mudawar (2003)	R multi, H	Copper + Lexan cover	0.349	–	Water	135–402	335
Sumith *et al.* (2003)	C single, VU	Stainless steel	1.45	–	Water	23–153	85
Yun *et al.* (2003)	C single, H	Stainless steel	6.0	Smooth	R134a, CO_2	170–340	182
Huo *et al.* (2004)	C single, VU	Stainless steel	2.01, 4.26	0.0009, 0.0004	R134a	100–500	365
Lee and Mudawar (2005)	R multi, H	Copper + Lexan cover	0.349	–	R134a	61–657	111
Saitoh *et al.* (2005)	C single, H	Stainless steel	0.51, 1.12, 3.1	Smooth	R134a	150, 300	420
Yun *et al.* (2005)	R multi, H	Stainless steel	1.14, 1.53, 1.54	–	CO_2	200–400	57
Muwanga and Hassan (2007)	C single, H	Stainless steel	1.067	–	FC72	770–1040	454
Zhao and Bansal (2007)	C single, H	Stainless steel	4.57	Smooth	CO_2	140–231	22
Agostini *et al.* (2008)	R multi, H	Silicon + Lexan cover	0.336	0.0005	R236fa	281–1370	593
Consolini (2008)	C single, H	Stainless steel	0.51, 0.79	0.0047, 0.0022	R134a, R236fa, R245fa	274–1435	650
Bertsch *et al.* (2009)	R multi, H	Copper + Lexan cover	0.544, 1.089	<0.0009, <0.0006	R134a, R245fa	19–336	332
In and Jeong (2009)	C single, H	Stainless steel	0.19	–	R123, R134a	314–470	256
Mastrullo *et al.* (2009)	C single, H	Stainless steel	6.0	Smooth	CO_2	200–349	143
Ohta *et al.* (2009)	C single, H	Stainless steel	0.51	–	FC72	107, 215	24
Wang *et al.* (2009)	C single, H	Stainless steel	1.3	–	R134a	321–836	365
Ducoulombier (2010)	C single, H	Stainless steel	0.529	0.0015–0.0030	CO_2	200–1400	1573
Hamdar *et al.* (2010)	R single, H	Aluminum	1.0	–	R152a	210–580	50
Martín-Callizo (2010)	C single, VU	Stainless steel	0.64	0.0012	R134a, R22	185–535	381
Ong (2010)	C single, H	Stainless steel	1.03, 2.20, 3.04,	0.0006, 0.0004, 0.0003	R134a, R236fa, R245fa	199–1608	2504
Tibiriçá and Ribatski (2010)	C single, H	Stainless steel	2.32	0.0001	R134a, R245fa	50–700	130
Ali *et al.* (2011)	C single, VU	Stainless steel	1.7	0.0001	R134a	75–600	152
Bang *et al.* (2011)	C single, H	Stainless steel	1.73	–	Water	100	65
Copetti *et al.* (2011)	C single, H	Stainless steel	2.62	0.0008	R134a	240–932	876
Mahmoud *et al.* (2011)	C single, VU	Stainless steel	1.1	0.0012	R134a	128–549	152
Oh and Son (2011a)	C single, H	Stainless steel	4.57	Smooth	CO_2	400–900	107
Oh and Son (2011b)	C single, H	Copper	1.77, 3.36, 5.35	Smooth	R134a, R22	200–500	153
Wu *et al.* (2011)	C single, H	Stainless steel	1.42	–	CO_2	300–600	419
Costa-Patry & John (2012)	R multi, H	Copper	0.295	–	R134a, R245fa, R1234ze	205–569	510
Karayiannis *et al.* (2012)	C single, VU	Stainless steel	1.1	0.0012	R134a	215–550	545
Li *et al.* (2012)	C single, H	Stainless steel	2.0	Smooth	R1234yf, R32	100–400	169
Tibiriçá *et al.* (2012)	C single, H	Stainless steel	1.0, 2.2	0.0006, 0.0004	R1234ze	300–600	30
Balasubramanian *et al.* (2013)	R multi, H	Copper	0.489, 0.504	0.00409, 0.00397	water	88–751	332
Davide Del Col *et al.* (2013)	C single, H	Copper	0.96	0.001354	R134a, R1234yf	200–600	93
Grauso *et al.* (2013)	C single, H	Stainless steel	6	Smooth	R1234ze(E), R134a	270.75–285.25	575
Vakili-Farahani *et al.* (2013)	R multi, VU	Aluminum	1.44	–	R245fa, R1234ze	100–400	138
Charnay *et al.* (2014)	C single, H	Stainless steel	3.0	–	R245fa	300–1500	285
Wang *et al.* (2014)	C single, H	Copper	6.0	Smooth	Propane	63.9–102.8	127
Anwar *et al.* (2015)	C single, VU	Stainless steel	1.6	0.000594	R1234yf	300–500	256
Charnay *et al.* (2015)	C single, H	Stainless steel	3.0	–	R245fa	300–1000	337
Markal *et al.* (2016)	R multi, H	Silicon	0.15	–	Water	51–92.6	20
Xu *et al.* (2016)	C single, H	Copper	0.501, 1.084, 2.0235	–	R134a	185–910	225
Sempértegui-Tapia & Ribatski (2017)	C multi, H	Stainless steel	0.868, 1.1	0.0026, 0.0097	R134a	200–800	685
Sempértegui-Tapia & Ribatski (2017)	C single, H	Stainless steel	1.1	0.0026	R134a, R600a, R1234yf, R1234ze	200–500	862
Fayyadh *et al.* (2017)	R multi, H	Copper	0.42	0.000716	R134a	50–300	50
Total							16,953

[a] C: circular, R: rectangular, H: horizontal, VU: vertical upward.

uncertainty to improve the model's ability to make accurate predictions. Machine learning methods have been used before to try and predict fluid characteristics, most commonly Artificial Neural Networks (ANNs) [25]. Such methods are useful because they are often able to obtain quicker and have a higher probability of generalization than large, numerical methods that may require numerous time and data. The trade-off is that sometimes the exact process of reaching the end result can be obscured within the layers of the ANN or other methods. That is to say that these methods do not necessarily follow traditional means to determine the solutions but instead learn the solutions using pattern analysis and learning from examples. Whereas two phase flow correlations would normally require an iterative procedure because of how fluid properties depend on temperature, ANNs only need input and output variables to train and input samples to test whether or not it recognized the correct patterns. Existing models usually use one or two hidden layers and a relatively low number of features [2].

Table 2
Summaries for all characteristic and numeric variables.

Characteristic Variables	Fluids	CO_2, FC72, Propane, R11, R113, R123, R1234ze, R1234yf, R134a, R152a, R22, R236fa, R245fa, R32, R600a, Water
	Channel Material	Stainless Steel, Aluminum, Brass, Copper, Copper + Lexan, Silicon, Silicon + Lexan
	Channel Geometry	Circular, Rectangular
	Channel Orientation	Horizontal, Vertical Up
	Channel Numbers	Single, Multi-channels
	Flow Boiling Characteristic	Convective Boiling, Nucleate Boiling
Numeric Variables		D_h, T, G, x, h, $q"$, α, R, P, P_C, ρ_g, ρ_f, h_{fg}, $C_{v,g}$, $C_{v,f}$, $C_{p,g}$, $C_{p,f}$, μ_g, μ_f, k_g, k_f, σ, PH, PF, Pr_g, Pr_f, Re_g, Re_f, Re_{go}, Re_{fo}, Bo, PR, X_{tt}, We_g, We_f, We_{go}, We_{fo}, Fr_g, Fr_f, Fr_{go}, Fr_{fo}, Co, Bd, Su_g, Su_f, Pe_g, Pe_f

1.4. Objectives of study

Qiu et al. [26] has amassed a database containing flow boiling heat transfer coefficient data from 50 different experiments using a total of 12 different fluids [27–76]. These experiments collected approximately 61 different features describing either fluid or channel properties. For the experiments conducted, the observed flow boiling behavior ranged from flow completely in the vapor region to completely in the liquid region and channels, ranging from 0.15 to 6.5 mm in hydraulic diameter. Table 1 contains important channel, fluid, and setup information about each experiment compiled into the database. This study will use a number of data science methods to try and determine a technique that can be employed to accurately predict the heat transfer coefficient during flow boiling in mini/micro-channels. Exploratory data science will be used to obtain confidence in the data and investigate relationships between feature variables before employing machine learning algorithms. Missing data was filled in using random forest nonparametric imputation. A variety of feature analysis techniques were employed to combine and select different optimal feature variables as input values before training any models including principal component analysis to

Table 3
Summaries of all fluid variables.

Variable	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.
D_h (mm)	0.150	0.529	1.1	1.681	2.2	6.5
T (K)	233.2	288.1	304.1	302.6	305.1	474.5
G (kg/m^{2}s)	19.45	299.80	400.00	463.44	568.00	1608.00
x	0.0001	0.1160	0.272	0.337	0.521	0.998
h (W/m^{2}K)	292	5490	8620	10,893	13,000	1,614,499
$q"$ (W/cm^2)	0.215	2.000	3.5	5.896	6.7	95.917
α	0.104	0.324	0.328	0.918	1.060	3.870
R	0.000	0.000	0.001	0.001	0.001	0.005
P (bar)	0.999	3.314	6.078	9.157	7.99	57.291
P_c (bar)	18.30	36.40	40.59	54.67	40.60	341.59
ρ_g(kg/m^3)	0.597	18.383	29.539	35.934	40.175	194.2
ρ_f(kg/m^3)	529.8	1084.5	1183.5	1162.7	1265.6	1593.4
h_{fg}(kJ/kg)	92.61	163.52	177.35	286.48	194.74	2256.44
$C_{v,g}$(kJ/kg/K)	0.568	0.8113	0.84	0.889	0.872	2.076
$C_{v,f}$(kJ/kg/K)	0.610	0.907	0.92	1.045	0.944	3.768
$C_{p,g}$(kJ/kg/K)	0.658	0.944	1.065	1.2	1.221	4.559
$C_{p,f}$(kJ/kg/K)	0.918	1.352	1.428	1.715	1.539	4.503
μ_g(μPa*s)	6.492	11.175	11.869	11.936	12.350	18.187
μ_f(μPa*s)	66.15	164.57	193.72	213.17	264.88	456.06
k_g(W/m/K)	0.00940	0.0134	0.0144	0.0153	0.0162	0.0403
k_f(W/m/K)	0.053	0.0742	0.0809	0.113	0.088	0.683
σ(N/m)	0.0012	0.00724	0.0083	0.0107	0.01	0.0589
PH(mm)	0.450	2.287	3.456	5.441	6.912	20.420
PF(mm)	0.596	2.481	3.455	5.498	6.911	20.420
Pr_g	0.670	0.812	0.879	0.938	0.937	6.7
Pr_f	0.906	3.228	3.414	3.692	4.093	8.914
Re_g	2.9	3504.4	10269.0	19250.1	24418.4	359794.7
Re_f	1.95	640.58	1429.94	2445.88	2888.41	55245.35
Re_{go}	623.5	21466.3	39099.2	59178.6	73641.0	410851.8
Re_{fo}	27.15	1176.75	2684.29	3819.89	4482.29	55,270
Bo	3.1×10^{-5}	2.4×10^{-4}	4.9×10^{-4}	7.3×10^{-4}	8.7×10^{-4}	2.2×10^{-2}
PR	0.00459	0.0861	0.150	0.177	0.202	0.776
X_{tt}	0.0006	0.197	0.466	1.637	1.147	590.779
We_g	0.000	8.707	52.410	231.062	218.998	20918.88
We_f	0.000	1.753	7.429	27.057	24.174	3139.78
We_{go}	1.37	333.64	763.65	1777.61	1905.44	32696.92
We_{fo}	0.0069	7.764	21.824	54.169	53.029	3142.589
Fr_g	0.0	197.6	1162.4	18957.2	4891.0	2899757.5
Fr_f	0.0000	0.941	4.167	15.883	14.215	434.5072
Fr_{go}	27	4720	15,153	2,220,191	53,536	328,588,796
Fr_{fo}	0.0225	3.117	10.585	32.458	30.750	439.105
Co	0.00088	0.148	0.319	0.839	0.714	240.050
Bd	0.00359	0.502	1.743	7.051	7.460	98.608
Su_g	35,116	1,072,748	2,030,903	2,875,745	4,029,987	14,043,660
Su_f	29,502	190,951	290,002	444,463	567,425	3,136,614
Pe_g	2.9	3013.3	9346.9	18392.7	23428.8	485744.1
Pe_f	7.18	2197.96	5165.76	8813.41	10814.26	181846.28

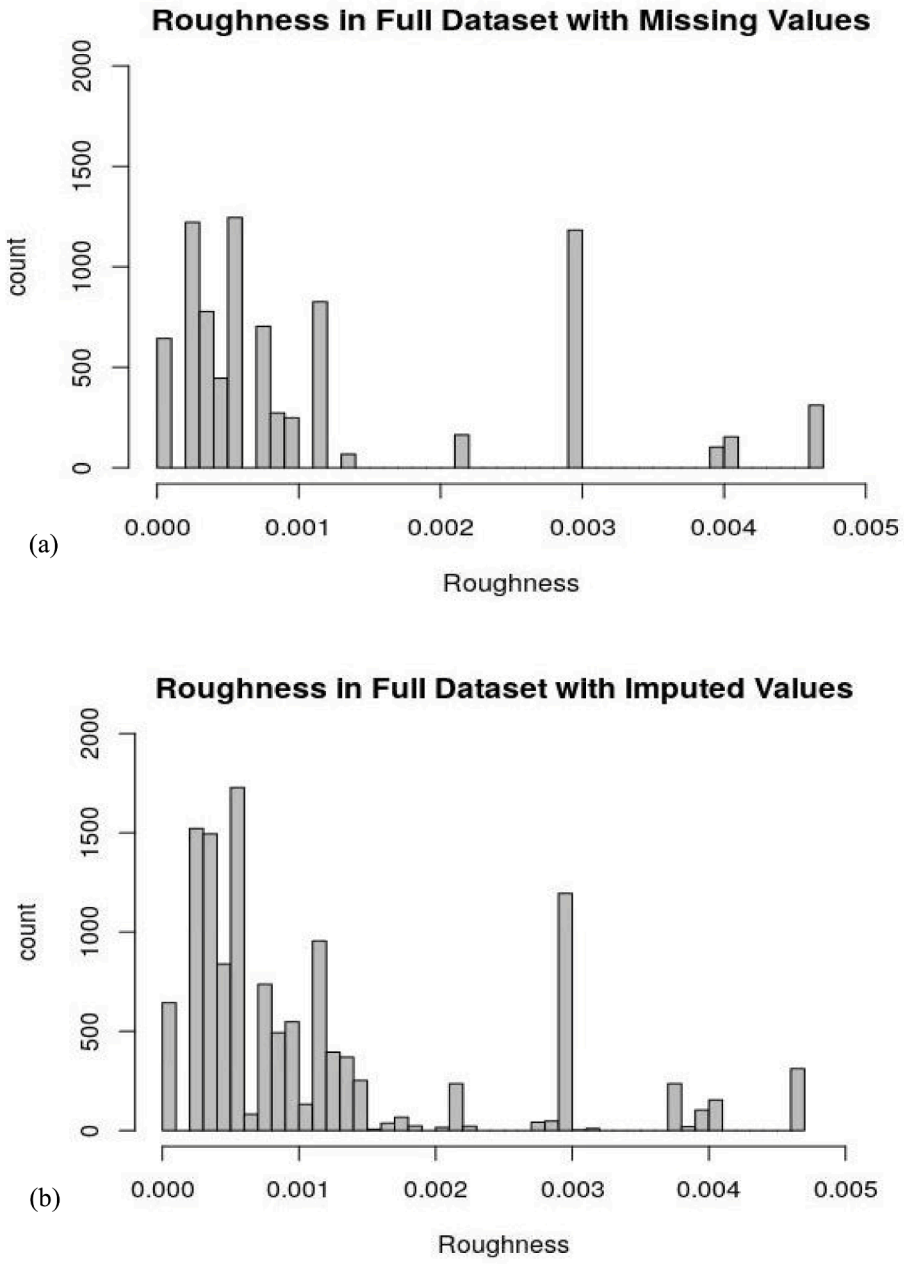

Fig. 3. The distribution of Roughness values with (a) values missing and (b) imputed values.

reduce the overall dimensionality of the dataset and the Boruta package, variable importance, LASSO regression, genetic algorithms, and step-wise selection to reduce the number of original variables used when modeling while preserving as much information as possible. Seven different types of modeling were then performed on the datasets including multiple linear regression modeling, generalized additive modeling, multivariate adaptive regression splines, random forest modeling, gradient boosting machines, support vector machines and feedforward neural networks. These models were compared to semi-empirical correlations and analyzed for significant feature contributions.

2. Exploratory data analysis

Using machine learning methods to try to predict the heat transfer coefficient during multiphase flow in mini/micro-channels requires a large database with enough observations to sufficiently represent the flow under a variety of possible conditions. If the data used to create the model does not adequately represent the flow conditions one may encounter when collecting data during two-phase flow conditions, the model will not be very accurate. It is important to analyze and assess the database being used to define its limits and the limits of any models created from it. Data exploration and analysis was done using Rstudio, an Integrated Development Environment (IDE) for R, a programming language primarily used for statistical analysis, data analytics, and graphics visualization [77,78].

When performing data analysis, it is important to have data used for training models and independent data used for testing. The datasets should be representative of similar populations but cannot contain any overlapping observations so that data used to verify the model can remain completely independent. The dataset was generated to contain a total of 53 variables that included characteristic and numeric variables as shown in Table 2.

The training data is 75 percent of the full dataset and is used to help

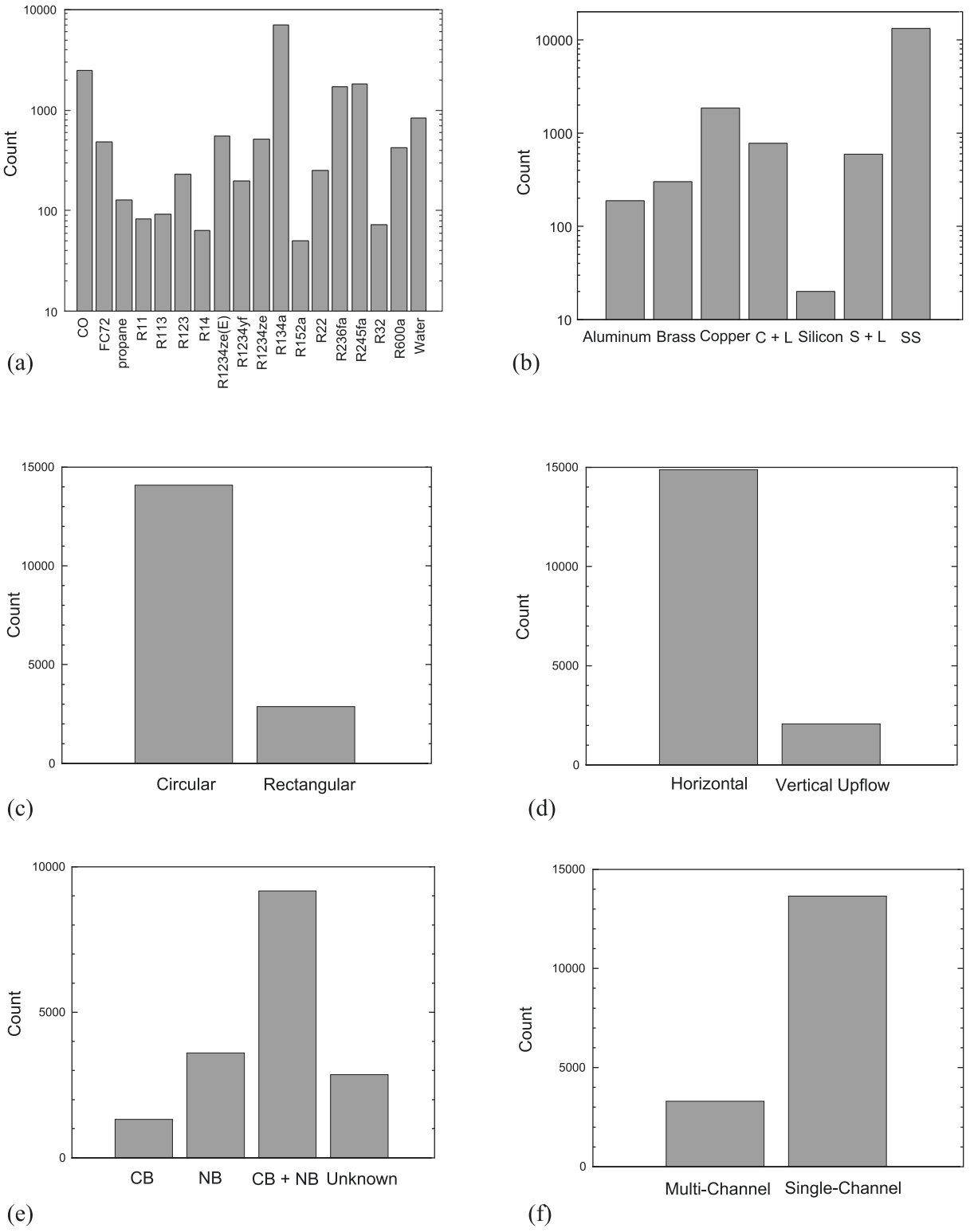

Fig. 4. Frequency Distributions of Categorical Flow Boiling Variables by (a) Fluid, (b) Channel Material, (c) Channel Geometry, (d) Channel Orientation, (e) Heat Transfer Characteristics, and (f) Number of Channels.

teach patterns to the model. The other 25 percent of the data is the testing data, which is meant to help understand how the model would perform with independent data collected in separate experiments. It helps determine performance and accuracy completely free of bias [77]. It's important that the data in both training and testing sets are proportionate and representative of the full dataset, so the data is divided into 75% and 25% sample sizes while keeping the proportions of each individual fluid data equal to their proportions in the full dataset. After aggregating and cleaning the data, it is important to get a large overview of the dataset and the values for each variable. The six variable summaries in Table 3 primarily illuminate the range of each numeric variable and the orders of magnitude for each variable compared to each other. For example, the Reynolds number ranges from an order of magnitude of one to an order of magnitude of 10^4, the heat transfer

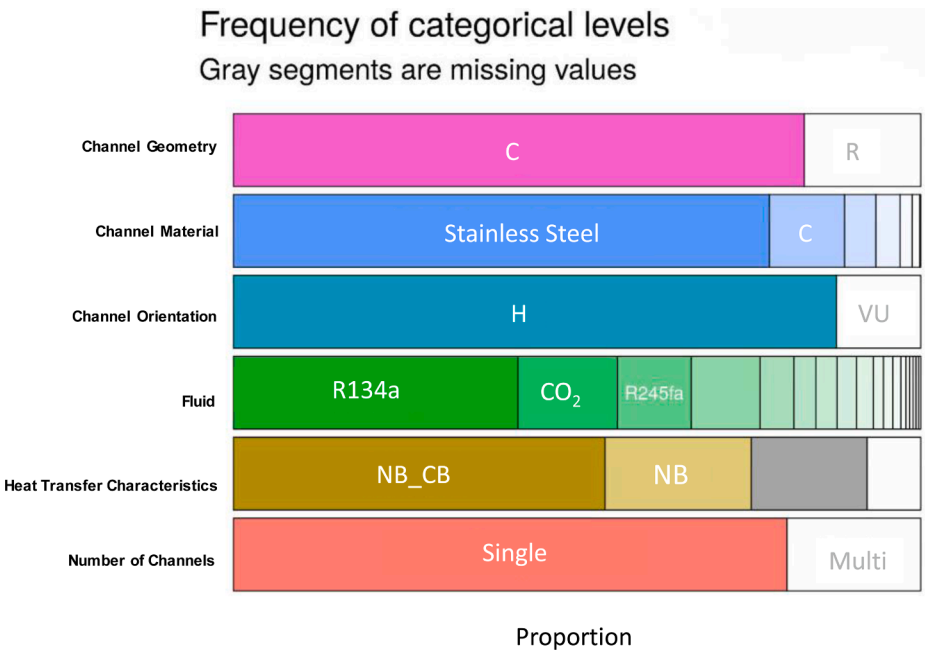

Fig. 5. Stacked Frequency Distributions of Categorical Flow Boiling Variables.

coefficient ranges from an order of magnitude of 10^2 to an order of magnitude of 10^6, and the Martinelli parameter ranges from an order of magnitude from 10^{-4} to an order of magnitude of 10^2. This informs us that it will be necessary to scale and center the variables. It is also important to get a profile of any missing data in the dataset.

With 16,953 observations, it is very difficult to know how much is missing just by glancing at the dataset. The heat transfer characteristics, Roughness, Aspect Ratio, $C_{v,g}$ and $C_{v,f}$ all have missing values. The channel aspect ratio, is a parameter relevant to only rectangular channels and is not needed for 80% of the observations, so the variable was completely removed from the dataset. The channel roughness is missing in approximately 34% of observations, and the heat transfer characteristics, $C_{v,g}$ and $C_{v,f}$ are missing in approximately 16% of observations. These percentages are much smaller and the variables may have some correlation with the heat transfer coefficient, so they were not excluded. Instead, they were left blank for any method that could be run with empty values and filled in using Random Forest Non-Parametric Imputation [79]. A random forest algorithm will make the same number of decision trees as their observations in the dataset, and then take a sample of the same size with replacement from that data set. Each tree is split with a random selection of possible variables. The size of this selection is determined by the user. The missing variables are then imputed using the average values of all trees. A random forest model can handle the 53 variables contained in the dataset and identify and report the most important contributing variables. It is also able to estimate the error on an equivalently sized testing dataset not included in the model, known as the Out of Bag (OOB) error. The imputed values have maintained the approximate distribution of the original dataset, but values have been added based on the other variables in the dataset. An example of this can be shown in Fig. 3(a) and (b). The OOB error rates for all missing values were all on the order of magnitude of 10^{-7}, so considering the percentage of missing values in each variable, and the expected contribution to the heat transfer coefficient, this imputed data is considered sufficient for future analysis, as there is a low probability of any results or conclusions being over reliant on the imputed data.

The data is very unevenly distributed as observed in Table 1. The majority of the data appears to be skewed right, meaning a large number of values occupy the lower part of the data range. For example, the heat transfer coefficient and dependent variable, h, has a median of approximately 8620 $W/m^2 \cdot K$ but a maximum value of 1,614,499/$m^2 \cdot K$, which appears a bit high. Similar results are observed in a number of other variables such as *We*, X_{tt}, *Fr*, *Co*, *Bd*, and *Re*. For these values, it appears 90% of the data is concentrated in the lowest two bin widths, or approximately the lowest 10% of the range of data, and the rest of the data is spread out as far as five standard deviations above the mean. Some of this could be caused by outliers, and/or experimental error, but more is revealed when looking at the bar frequency charts for categorical factor data in Fig. 4. Approximately 7000 observations, over 40% of the total data, are from experiments using R134a. Approximately 13,000 observations, over 75% of the total data, are explained by only 4 of the 12 different fluids. With this irregular distribution of data, it is more likely that the results will demonstrate a network that can predict the heat transfer coefficient for a specific subset of conditions than for the complete range of conditions represented in the database. When the fluid, channel material, channel orientation, heat transfer characteristics, channel geometry, and number of channels are all largely skewed in favor of one value, it creates unbalanced results and insignificant information for underrepresented conditions [80]. This works against the training effects of those conditions and makes validation and testing more difficult. Approximately 75% of the data comes from experiments that use circular channels rather than rectangular channels and whose channels are made out of stainless steel. Over 85% of the data comes from experiments that use a single channel over multiple and place the channel in a horizontal orientation. By combining this information into a stacked bar plot, we can achieve a better visualization of how the imbalanced data is distributed in Fig. 5.

An ideal database would contain equal proportions within each factor variable or enough data so that each factor has a representation of more than at least 10% of the total number of samples to make sure there is a big enough sample size. The channel material and fluid variables specifically have some factors with extremely low numbers of observations, such as R152a, which has 38 out of the 16,953 observations in the dataset.

After further exploring many of these variables with extremely uneven distributions, there are three primary correlations for a majority of these outlying variables. The first is that a majority of these outliers come from experiments using water as the working fluid, especially from the Balasubramanian et al. experiment [64]. It's unclear what specifically about this experiment causes significantly higher values for most

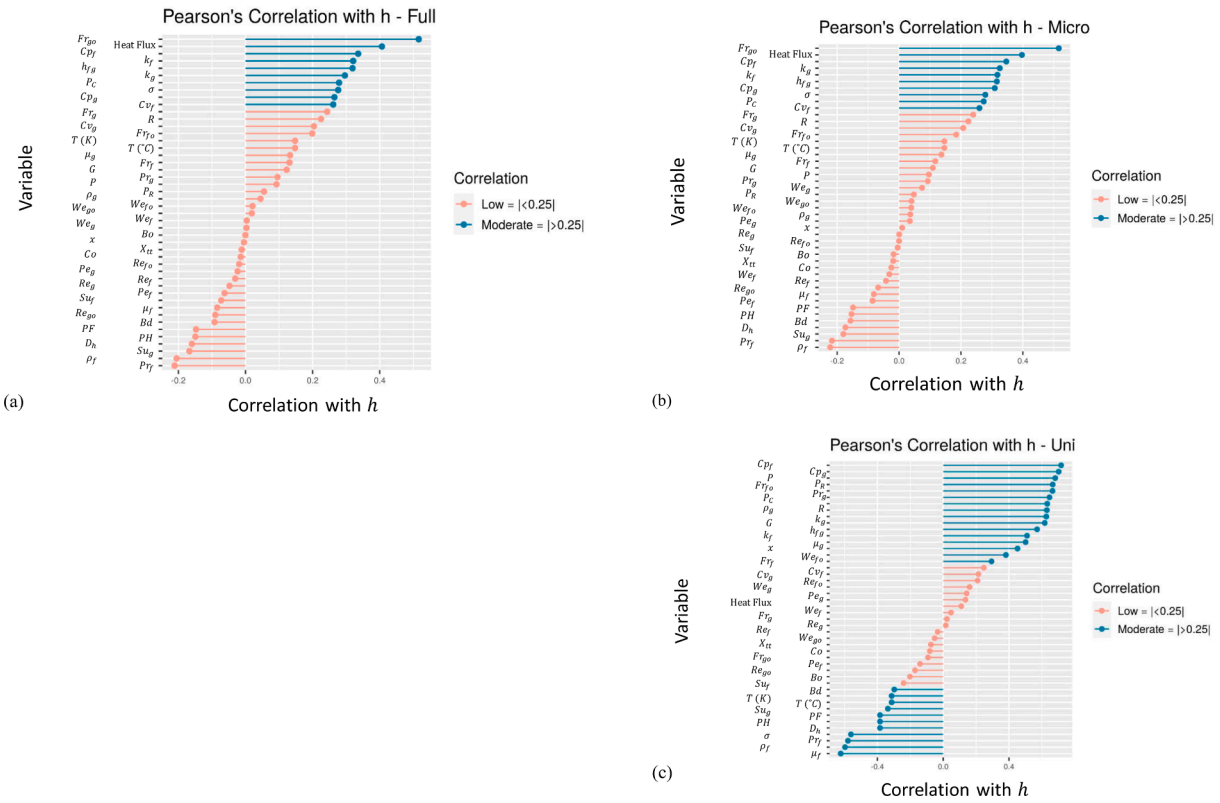

Fig. 6. Pearson's Correlation vs. h for (a) Full, (b) Microchannel, and (c) Unicategorical datasets.

of the variables, but one theory is that the experiment uses a significantly higher magnitude of mass velocity (500–700 $kg/m^2 \cdot s$) compared to other experiments using water at much lower magnitudes of mass velocities (80–150 $kg/m^2 \cdot s$). The final major correlation is that a significant portion of the outliers were from experiments conducted using channels with diameters greater than 3 mm. There is no definitive classification system between microchannels, minichannels, and conventional channels, however, one of the most commonly accepted classification systems found in existing research differentiates these types of channels by using a channel diameter of 3 mm as the threshold that separates a microchannel from a minichannel [81]. It is possible that the fluid properties in experiments where the channel diameter is greater than 3 mm mimic minichannel or conventional channel flow as opposed to microchannel flow.

As a result of this further explanation, all additional methods will be conducted on the full dataset and two data subsets to determine if these outliers have a significant impact on the ability to predict the heat transfer coefficient. The first subset of data, as previously mentioned, will contain all observations from the largest single categorical subset of data. Looking at Fig. 5, it can be seen that this subset consists of a single circular channel made out of stainless steel in a horizontal orientation with both nucleate and convection boiling characteristics. All relevant fluids used in experiments with the aforementioned channel type were used. The second subset will contain all observations from experiments with channel diameters of$<$3 mm. The single categorical dataset contains 7245 of the original 16,953 observations and will be referred to as the unicategorical dataset, and the microchannel dataset contains 14,169 observations.

Correlation plots can provide a look at how variables are related. The Pearson's correlation plots in Fig. 6 (a) can show how all independent variables in the full dataset are correlated with *h*. The heat transfer coefficient does not have any strong positive or negative correlations with a single variable beyond a measure of 0.5, so there are not obvious variables to look to include when performing feature selection. The results vary from −0.22 to 0.52, which is a moderately weak negative correlation to a moderate positive correlation. Looking at the correlation values for the microchannel dataset as shown in Fig. 6(b), we see an extremely similar trend. The overall range of correlations is the same, and the absolute value of the strength of most of the variables is also within 0.03–0.05 of the full dataset, but some values with a weak negative correlation shift to a weak positive correlation and vice versa. For the unicategorical dataset as shown in Fig. 6(c), however, the correlations appear to change significantly. There are now 10 variables with a positive correlation greater than 0.5 as opposed to 2, and the upper limit of the positive correlation values approaches 0.56 instead of 0.52. Additionally, the lower limit of the negative correlation values approaches −0.52 and there are 11 values with a stronger negative correlation and the lower limit of the full dataset, −0.22. This indicates that it may be easier for a machine learning algorithm to predict the heat transfer coefficient for a specific set of variables for a specific case of categorical variables as opposed to for the whole dataset. This information is particularly useful when it comes to dimensionality reduction to narrow the total number of input variables. Every additional input that a machine learning algorithm incorporates significantly increases the total processing time and complexity.

3. Feature selection

Dimensionality reduction aims at reducing the number of independent variables by eliminating those that do not have a significant impact on *h* or performing linear combinations to turn existing variables into new variables. Feature selection is a process for calculating and choosing the features that contribute most to the dependent variable and removing the irrelevant or extraneous features that can decrease the accuracy of the model. Feature selection is a common method of reducing the number of variables, however, most feature selection methods include a manual component of selecting an arbitrary contribution value, below which the variables are deemed irrelevant [82].

Other variable reducing methods such as Principal Component Analysis (PCA) create new variables from linear combinations of existing variables to reduce complexity and dimensionality while minimizing the loss of information. Dimensionality reduction using PCA along with five different feature selection methods were performed to get a large range of different feature sizes and combinations.

3.1. Principal component analysis

PCA is a technique that treats each unique variable as a dimension and attempts to project all of the data onto a lower set of dimensions using linear combinations and removes all of the irrelevant variables while maintaining as much of the information as possible [83]. It is an unsupervised learning method, meaning the dependent variable is removed before the linear combinations are determined. Each of these individual linear combinations are known as principal components. If there are n predictors numbered $p_1, p_2, \ldots, p_n$, there can exist m relevant principal components where $Z^j = \Phi^j_1 p_1 + \Phi^j_2 p_2 + \ldots + \Phi^j_n p_n$. Here, Φ is a weighting factor based on the average squared euclidean distance between each predictor, p and placed under a constraint such that the sum of the squares of the loading factors is equal to one. The first principal component creates a line in a space of p dimensions closest to the n observations in the direction of the greatest variance in the dataset. The first principal component will capture the largest amount of information and variability in the data. The second principal component follows the same formula in a direction orthogonal to the first principal component while also capturing the greatest amount of the remaining variance that is possible. Therefore, there should be zero correlation between any of the remaining principal components. All successive principal components will follow the same trend, capturing the greatest amount of remaining variance with zero correlation with the previous components [84].

3.2. Boruta

The first algorithm employed for feature selection was the Boruta package in R [85]. Boruta utilizes random forest algorithms, similar to those used in variable imputation, to assess the importance of each feature and keep only those that surpass a threshold. Typically, the thresholds used in this type of assessment have a level of arbitrariness to them. They are usually assigned a value consistent with common practice or that captures a level of variance or importance that is acceptable to the user. Boruta chooses the threshold on its own using shadow features. Before starting the random forests, Boruta duplicates the entered dataset but randomly shuffles the rows within each feature, resulting in a permuted "shadow dataset" attached to the original one [86]. Then a random forest model is fitted on the combined data frame. This model creates a large number of decision trees, which create partitions in the data in an attempt to generate a set of variables that capture the greatest overall contribution to h. The more accuracy lost due to a random partition of values within a variable, the greater the importance of the variable [87]. Once a single iteration of the random forest is complete, the Z-score for the accuracy loss is calculated across all individual decision trees for each feature. This Z-score is compared with the Z-scored for the equivalent feature in the "shadow dataset." If the Z-score is greater, i.e., if the average accuracy loss across all decision trees divided by the standard deviation is higher than the maximum Z-score of any completely randomized feature, that feature is judged to be an important contributor in the dataset. This algorithm dictates that a feature is a significant contributor to an individual model if the average loss in accuracy of a model due to a random partition with respect to a given feature is greater than the average loss in accuracy of a model due to a random partition with respect to its randomized counterpart [87]. Boruta performs 20 or 100 iterations of this procedure while treating each variable's significance or insignificance as a binary outcome of a series of trials. This creates a binomial distribution that Boruta uses to determine whether or not the variable is rejected, accepted, or tentative depending on whether or not the p-value of the number of confirmations in 100 trials of a binomial distribution is significantly greater than, significantly less than, or around the value of 0.05. This provides a statistical benchmark for feature selection in both classification and regression problems while still accounting for multivariate correlations and dependencies. This is a computationally expensive algorithm with a long runtime of approximately 2 h on the same GPU node with 2 cores and 36 GB of memory, but it has been shown to produce results that are reliable when analyzed alongside other feature selection methods [86].

3.3. Recursive partitioning

Recursive partitioning, or the recursive partitioning method, works by creating a decision tree to split the dependent variable into groups as homogeneous as possible. Each split is made without consideration for how effective the split will be in the future, meaning it focuses on local optimization and may not locate the universally best possible decision tree [88]. The tree stops splitting when doing so no longer reduces the mean squared error when predicting the dependent variable. This splitting could potentially continue until each unique value of the heat transfer coefficient, a continuous numerical variable in the dataset was partitioned, but that would significantly increase the chance of overfitting any future models. In order to reduce the decision tree to a model that appropriately fits the dataset, recursive partitioning adds a penalty proportional to the number of branches in the decision tree while minimizing the variance of the dependent variable. The variable importance metric is the sum of the reduction in the mean squared error at each branch of the tree [89].

3.4. Recursive feature elimination

Recursive Feature Elimination (RFE) builds a random forest or other specified model using all variables in the dataset, calculates the importance for each variable using the reduction in the mean squared error at each branch of the tree added together for each variable, similar to recursive partitioning, ranks the variables by importance, removes the worst variables, builds another model, and repeats the process until the pre-specified number of features remain [90]. Each individual repetition of this process is known as backwards selection. For each dataset, RFE algorithms were performed with 10, 20, 30, 40, 50, 60, and all 81 features.

3.5. LASSO

LASSO stands for Least Absolute Shrinkage and Selection Operator and works by building a regression model with all variables and then penalizing the variables with the smallest coefficients by driving them to zero to reduce overfitting. That is, the values of scaled variables that have a smaller weighted coefficient in a linear model contribute less to the overall model and are driven to zero by LASSO [91]. First the dataset is trained on a generalized linear model. A generalized linear model is a method that models the dependent variable using a linear combination of every feature in the dataset with an intercept as a systematic component and adds two additional components. The first is a link function that uses the linear model to explain the mean of the dependent variable. The second is a variance function that uses the mean to calculate the variance and probability distribution of the dependent variable to try and estimate its randomness. Then the weight of the tuning parameter that drives the smallest coefficients to zero is calculated by finding the value that produces the lowest mean squared error in the model [92]. All of the remaining nonzero coefficients will be used when modeling.

3.6. Stepwise selection

The stepwise selection method works by training first two linear models: one with no features included in the training model, only an intercept, and one with every feature included when training the model. The method then employs an algorithm which iteratively and systematically adds and subtracts variables from the two models until it reaches the model with the lowest Akaike Information Criterion (AIC) score. An AIC score is a criterion that evaluates all models in a method relative to each other and outputting a value based on the model's fit to the training data and its complexity. A more complex model incurs a higher penalty, and a lower AIC score signifies a better model. The model's fit to the training data is measured using likelihood estimation which takes the logarithm of the probability of seeing the dependent variable values in the training data as an output of the constructed model. The best model is the one that achieves the best balance of increasing the number of features to have a higher likelihood of achieving outputs closest to the values of the heat transfer coefficient while keeping the model's complexity as low as possible. Table A.1 summarizes the amount of and which features were selected using each method for each dataset.

4. Modeling analysis

Feature selection provided various combinations of independent variables that could be used to predict the heat transfer coefficient. This allowed us to test various types of models. A total of seven different types of models were used to test the six different subsets of variables for each of the three datasets for a total of 126 individual models. These types of models are: multiple linear regression models, generalized additive models (GAMs), multivariate adaptive regression splines (MARS), random forest models, gradient boosted machines (GBMs), support vector machines (SVMs), and artificial neural networks (ANNs).

The primary measure of accuracy used in comparing and analyzing these models will be the Mean Absolute Percentage Error (MAPE), defined as the absolute value of the ratio of the difference between the observed and predicted values divided by the observed value summed for every observation and divided by the total number of observations i.e.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{X_{obs,i} - X_{pred,i}}{X_{obs,i}}\right| * 100 \tag{1}$$

This value is used as a primary basis of accuracy for a number of reasons [93]. The experimental heat transfer coefficients range from 292 to 1,614,499 W/m^2•K meaning that a divide by zero or approximate situation resulting in an infinite error is unlikely to occur. The mean heat transfer coefficient is approximately 11,000, so any RMSE or MAPE used in absolute terms will be fairly large. It is easier to comprehend the error as a percentage of the average value of the heat transfer coefficient when comparing models. The RMSE will still be reported as a secondary measure of accuracy and is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{obs,i} - X_{pred,i}\right)^2}{n}} \tag{2}$$

4.1. Multiple linear regression models

Multiple linear regression modeling is the practice of building linear models explaining a single, continuous, dependent variable using multiple, continuous or categorical, independent variables. Essentially it transforms the simple linear relationship $y = mx + B$, where y is the dependent variable, m is the slope of the line, x is the independent variable, and B is the intercept, into a relationship of linear combinations that can be described as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{3}$$

Here, β is a numerical coefficient and x is the value for a dependent variable amongst a total of n dependent variables [94]. Each variable is only included at the first degree to form a complex linear relationship. Many of the models were manually tuned afterwards by selectively removing features and performing transformations. For example, it was observed that applying a logarithmic or square root transformation to the dependent variable dramatically increased the accuracy of the linear models.

4.2. Generalized additive models (GAMs)

Generalized additive models transform the standard equation modeling the response as a combination of linear terms into an equation modeling the response as a combination of functions in the form of

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) \tag{4}$$

Each $f(x)$ is its own unique function. This method also maintains the additive features of linear models in that each term is specified. GAMs assist in eliminating the necessity of deciding the order of a polynomial and make models more equipped to deal with high rates of change, especially at the extreme ends of the model [95]. These functions, often called splines or smooth functions, are summed expansions of simpler, weighted basis functions evaluated at the values of each variable. The number of basis functions is determined by specifying knots, which are specific variables that the user desires to be nonlinear, and the basis function is only nonzero for the knot in which it was defined. The default basis function for the package used to make these GAMs is thin-plate spline function, which is a basis function developed after a physical model with a closed solution and few necessary tuning steps but that often comes at a higher computational cost. This model is typically seen as

$$g(y) = \beta_0 + f(x) + \lambda \tag{5}$$

Here, β_0 is a simple intercept, $f(x)$ is a function to the n^{th} degree where n is often determined for each knot to generalize cross-validation, and λ is an error term that smooths the data based on a penalized least squares method often likened to the amount of pressure needed to bend a thin sheet of metal between variable values x_1 and x_2 [95]. It is a penalized term used to avoid overfitting. This is a useful basis function for the fluid dataset because we have very little prior knowledge as to the relationships between variables and can therefore not easily determine where knots should be placed and/or to what degree [95].

4.3. Multivariate adaptive regression splines (MARS)

MARS models have a similar foundation of basis functions to GAMS but often act as an extension of GAMS because they do not make the assumption that the weighted coefficients of independent variables are uniform across their range of values. MARS models are also built using an iterative method beginning with one of two possible values for basis functions. The first is a constant value of 1, signifying an intercept. The second is a characteristic hinge function taking the form $\max(0, x - c)$ where x is the value of an independent variable, and c is a constant representing the location of a knot that separates two piecewise linear functions [96,97]. The basis function can also be the product of two or more hinge functions that can model interactions between multiple independent variables. A single independent variable can include multiple knots for every instance the linear function changes. MARS models are then built by creating independent linear regression models for each independent variable and dividing those models across the entire range of the independent variable based on where the slope of the line changes, and the exact value of the hinge function is determined by the maximum reduction in the sum-of-squares residual error [96]. In order to counteract overfitting, the MARS model then constructs a least-squares model to prune itself by iteratively removing the least influential knots until the best configuration of knots for each variable is

determined these configurations are then compared multiple times using generalized cross validation [97]. The user can tune this process by specifying the degree of interactions at the knots and the number of terms to keep during the pruning process.

4.4. Random forest models

Random Forests use an ensemble of decisions trees to create partitions in the dataset based on which variables can best predict the heat transfer coefficient [98]. A process known as bagging limits the variance of individual decision trees, but relationships between the independent and dependent variables often result in extremely similar tree composition and dependencies between trees. Additional randomness is introduced to the dataset using bootstrap resampling while only considering some of the variables at each partition within a given decision tree.

4.5. Gradient boosted machines

Gradient Boosted Machines (GBMs) rely on an extensive network (often thousands) of decision trees with a very limited number of partitions that are dependent and build off one another to learn and, in aggregate, produce models with high predictive accuracy. Rather than averaging the predictions of independent decision trees, GBMs first select a weak learning model as its base-learning model. Typically, and in this case, the weak learning model consists of generating shallow decision trees with low computational cost. These shallow trees iterate and sequentially learn from one another, gradually correcting the elements of the weak model with the most errors in future iterations while avoiding overfitting thanks to only making small improvements at a time and using cross validation to limit overfitting throughout the process. Similar to GAMs, each successive decision tree is additive, and the overall GBM model can be described using the equation $GBM(x) = \sum_{n=1}^{T} f^n(x)$ where *GBM* is the overall model, T is the total number of decision trees, and f is each individual decision tree [98]. Every subsequent decision tree is fitted to the residuals of the previous decision tree until overfitting is detected via cross validation.

4.6. Support vector machines

Support vector machines (SVMs) are a machine learning algorithm typically used for classification because their function is to analyze data across multiple dimensions and determine the optimal way to separate data into different classes. It performs this analysis by transforming the data into a higher level of dimensions using a specified kernel function that uses a predetermined formula, often based on known statistical nonlinear optimization techniques, to transform the data into a higher dimensional space before trying to separate the observations [98]. The "line" or separator used to differentiate observations in higher dimensional space or to predict a continuous dependent variable for a regression problem is called a hyperplane. When conducting support vector regression, the objective is to define two decision boundaries a fixed distance on either side of the hyperplane. This distance is chosen based on the data points that lie closest to the hyperplane. This technique is very effective for datasets with many variables because the kernel function, while heavily influential to the outcome of the model, can effectively transform and enlarge the feature space [99]. Support vector regression models also do not dependent on the distribution of the features or dependent variable in the dataset, which is useful as the features in the fluid dataset are not normally distributed. The outcome of the support vector regression model is instead heavily dependent on the choice of the kernel function that is typically selected via tuning.

Table 4
Optimal neural network tuning parameters.

Parameter	Value
Layers	1 Input, 8 Hidden, 1 Output
Input Layer Size	500 Units
Hidden Layer Size	105, 75, 70, 60, 50, 30, 20, 10 Units
Output Layer Size	1 Unit
Activation Function	ReLU
L_2 Regularization	0.001
Batch Normalization	Employed After Each Hidden Layer
Optimizer	Adam
Learning Rate	0.001
β_1 Decay Rate	0.9
β_2 Decay Rate	0.999
Loss Metric	MSE
Validation Metric	MAPE
Epochs	400 (Early stopping at 30 consecutive epochs with no loss reduction) (Learning rate divided by 10 at 10 consecutive epochs with no loss reduction)
Batch Size	200
Validation Split	0.2

4.7. Neural networks

Artificial neural networks (ANNs) are a class of machine learning methods inspired by the brain. The algorithm accepts inputs that simulate the function of biological dendrites absorbing information into a neuron. A transfer function, net input, and activation function act as a biological cell body that communicates a generated response signal only if a sufficient amount of information, a metric determined by each individual neuron, is absorbed via the dendrites. There are endless combinations of constructions that can form a neural network. Typically, basic neural networks begin with an input layer that reads user-entered values, a hidden layer that conducts a majority of the learning, and an output layer that displays the results. The network typically reads the data and converts it into m input vectors each labeled $\vec{x_i}$ [100]. The connections are then mapped between neurons and labeled using weights that describe the neuron the information is coming from, i, the neuron the information is going to, j, and what two layers this connection lies between, k. These labels take the form $w_{ij}^{(k)}$. When the data is initially fed into a simple model, the weights often start out assigned at random. Hidden layers can also take in additional inputs known as bias terms which are constants, each with their own weight, that describe how the chunk of information contained within a given neuron compares to the average of the overall data. The bias terms and weighted inputs in a given neuron are added together and assessed to see if they contain sufficient information using an activation function Θ. A neuron's output to the subsequent layer, therefore, can be described as $y_j = \Theta(b_j) + \mathrm{k}$ [98]. After all of the outputs are determined, a performance function P is used to compare the actual output values to the predicted ones. This function is usually derived from a user-specified metric. For the models being conducted on the fluid dataset, the function used was the mean squared error (MSE) defined as $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. The neural network iteratively adjusts the weights using the function $\Delta\vec{w} = r^*\left(\frac{\delta P}{\delta w_0}, \frac{\delta P}{\delta w_1}, \ldots, \frac{\delta P}{\delta w_n}\right)$ where r is a learning rate that essentially specifies a step size for the calculation [99]. Then using back-propagation, the weights are updated with the new values, and the process can be iterated until the MSE converges at a specified value or until a pre-specified number of iterations occur. The hidden layers then connect to an output layer that forms the results into n output vectors each labeled $\vec{y_i}$. As the dataset becomes more complex and the number of hidden layers and mapped connections needed increases, the amount of computation grows exponentially. The feedforward deep neural network used on the fluid dataset uses many hidden layers that can

Table 5
Previous universal saturated flow boiling heat transfer correlations.

Author(s)	Equation	Remarks
Bertsch *et al.* [101]	$h_{tp} = h_{nb} \bullet S + h_{cb} \bullet F$, $S = 1 - x$, $F = 1 + 80(x^2 - x^6)e^{-0.6C}$, $C = \sqrt{\frac{\sigma}{g(\rho_f - \rho_g)D_h^2}}$ $h_{nb} = 55P_R^{0.12}(-log_{10}P_R)^{-0.55}M^{-0.5}q''^{0.67}_H$, $h_{cb} = h_{cb,fo}(1-x) + h_{cb,go}x$ $h_{cb,fo} = \left(3.66 + \frac{0.0668\frac{D_h}{L}Re_{fo}Pr_f}{1+0.04\left(\frac{D_h}{L}Re_{fo}Pr_f\right)^{2/3}}\right)\frac{k_f}{D_h}$, $h_{cb,go} = \left(3.66 + \frac{0.0668\frac{D_h}{L}Re_{go}Pr_g}{1+0.04\left(\frac{D_h}{L}Re_{go}Pr_g\right)^{2/3}}\right)\frac{k_g}{D_h}$,	$D_h = 0.16 - 2.92$ mm, $G = 20 - 3000$ kg/m^{2}s Working fluid: water, nitrogen, methanol, pentane, heptane, benzene, FC-77, R11, R113, R12, R123, R134a, R141b, R236fa, R245fa, R410A 3,899 data points
Kim & Mudawar [102]	$h_{tp} = \left(h_{nb}^2 + h_{cb}^2\right)^{0.5}$, $h_{nb} = \left[2345\left(Bo\frac{P_H}{P_F}\right)^{0.7}P_R^{0.38}(1-x)^{-0.51}\right]\left(0.023Re_f^{0.8}Pr_f^{0.4}\right)\frac{k_f}{D_h}$, $h_{cb} = \left[5.2\left(Bo\frac{P_H}{P_F}\right)^{0.08}We_{fo}^{-0.52}(1-x)^{-0.51} + 3.5\left(\frac{1}{Xtt}\right)^{0.94}\left(\frac{\rho_g}{\rho_f}\right)^{0.25}\right]\left(0.023Re_f^{0.8}Pr_f^{0.4}\right)\frac{k_f}{D_h}$,	$D_h = 0.19 - 6.5$ mm, $G = 19 - 1608$ kg/m^{2}s, $Re_{fo} = 57 - 49{,}820$, $P_R = 0.005 - 0.69$ Working fluid: water, CO_2, FC-72, R11, R113, R123, R1234yf, R1234ze, R134a, R152a, R22, R236fa, R245fa, R32, R404A, R407C, R410A, R417A 10,805 data points

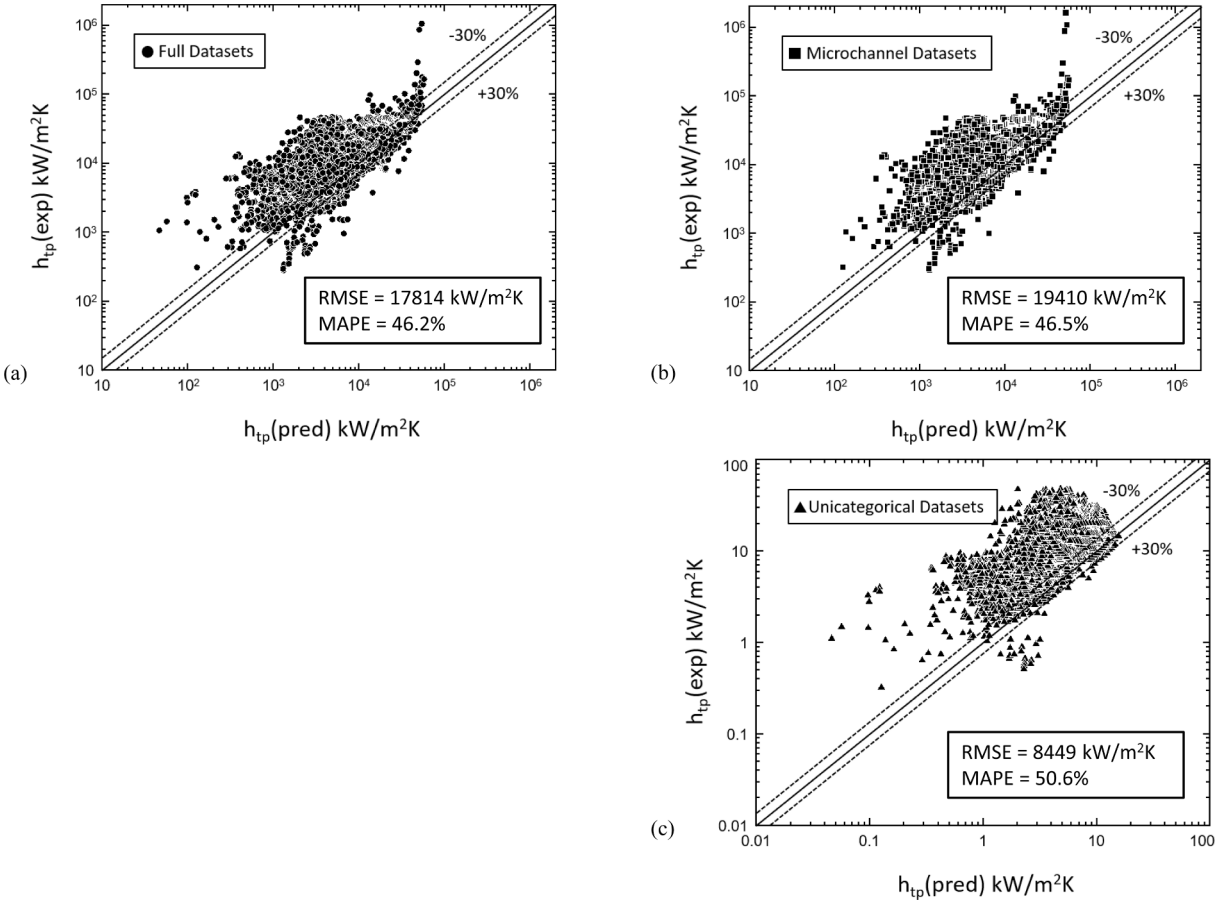

Fig. 7. Plots showing the Actual vs. Predicted Heat Transfer Coefficients using the correlation from Bertsch et al. [95] for the (a) Full, (b) Microchannel, and (c) Unicategorical datasets.

better model the complex and nonlinear relationships of the experimental conditions and fluid properties to the heat transfer coefficient.

More so than any of the previous models performed on this dataset, neural networks contain the greatest number of parameters that need to be predetermined by the user. It is unlikely that the first neural network parameters selected by the user is also the one that performs best on training dataset and is best able to be generalized to new data when run using the testing dataset. If the neural network is unable to match the training dataset very well, it is underfitting the data, and if the testing error is significantly higher than the training error, the neural network is likely overfitting the training dataset. The optimal neural network tuning parameters are shown in Table 4.

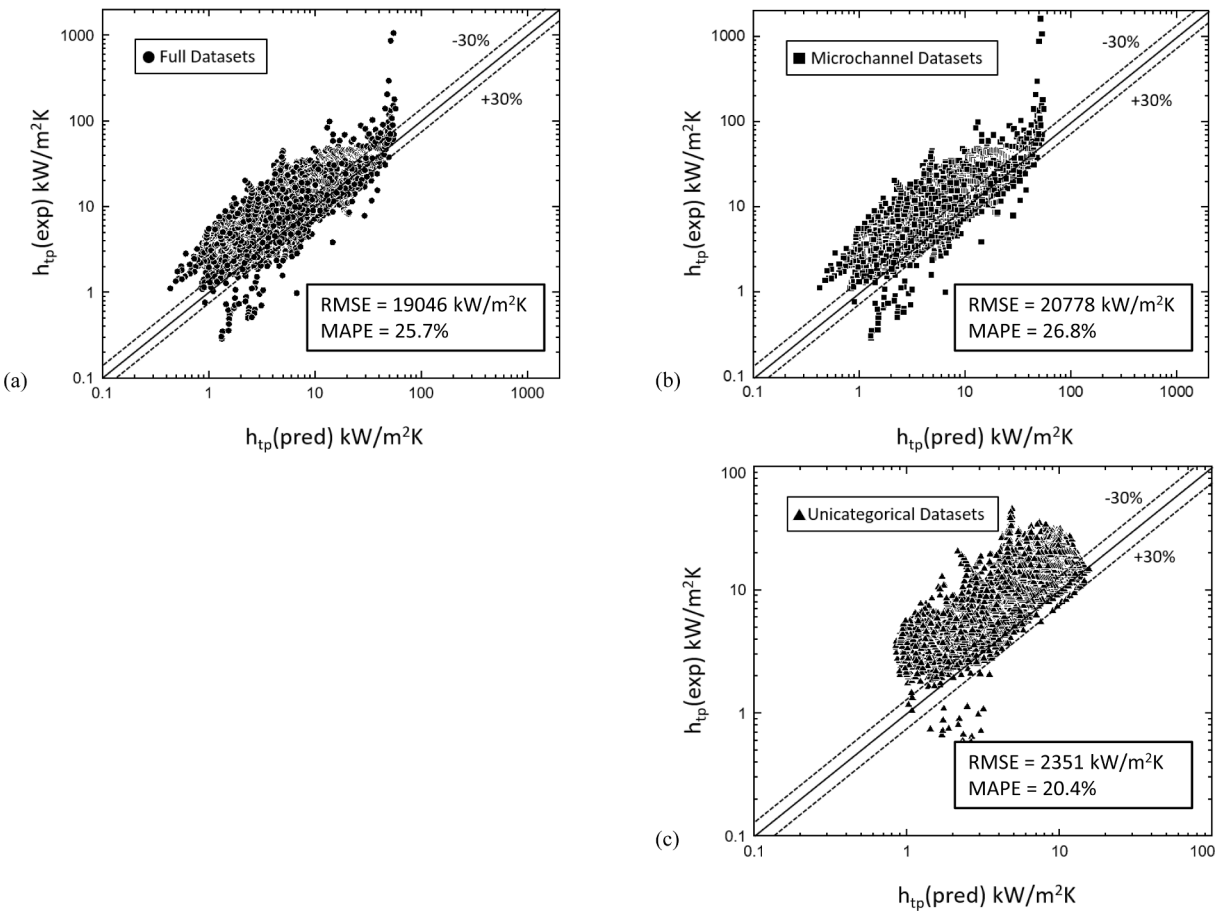

Fig. 8. Plots showing the Actual vs. Predicted Heat Transfer Coefficients using the correlation from Kim and Mudawar [96] for the (a) Full, (b) Microchannel, and (c) Unicategorical datasets.

5. Results and discussion

5.1. Semi-empirical correlations

Seven different types of models were performed using six different subsets of features on three different subsets of data for a total of 126 models. The features are scaled to a median of 0 and a standard deviation of 1 in order to make sure every feature is initially weighted equally. These models are compared with each other as well as two semi-empirical correlations that are used as a baseline. The two primary universal correlations summarized in Table 5, Berstch et al. [101] and Kim and Mudawar [102], used databases containing 389 and 10,805 observations respectively, while the fluid database being used here contains 16,953 observations. Additionally, there are experiments with fluids contained in this database, such as propane or FC72, that are not used in some or all of the previous correlations. We must also note that there are limitations to all of these previous correlations. The correlation from Berstch et al. [101] requires information regarding the length of the channel, which is not always provided in the experiments. The correlation from Kim and Mudawar [102] is only valid for data that meets the pre-dryout criterion determined by the dryout incipience quality, another numerically derived correlation that requires additional calculations and excludes observations beyond a certain threshold.

Nevertheless, these previous correlations provide a well-founded baseline for comparison when building models for which to predict the heat transfer coefficient. Heat transfer coefficients were calculated using the properties measured in the fluid database and compared with the experimentally measured heat transfer coefficients.

The Bertsch et al. [101] correlation uses approximately 20% of the data points of our full dataset and does not include many of the fluids or experimental conditions, and therefore has a relatively high MAPE of 46.2% for the full dataset, 46.5% for the microchannel dataset, and 50.6% for the unicategorical dataset as shown in Fig. 7(a), (b), and (c) respectively. The Kim and Mudawar [102] correlation is only valid for predryout data which, for the full dataset, amounted to 13,479 of the original 16,953 observations. Nevertheless, it performed more accurately than the Bertsch correlation with an MAPE of 25.7% for the full dataset, 26.8% for the microchannel dataset, and 20.4% percent for the unicategorical dataset as shown in Fig. 8(a), (b), and (c) respectively, and serves as the best point of comparison for future models. With these metrics able to be used for comparison, we can begin to compare the results of the models built using the fluid database.

5.2. Model performance

The full results of the models including the type of model, feature selection method used, dataset, MAPE, and RMSE are included in Table 6. The support vector machine models in Fig. 9(a), (b), and (c) produced the best overall models across the full, microchannel, and unicategorical datasets respectively. This may be due to the significantly fewer tuning parameters necessary for support vector machines. Support vector machines are only defined by the kernel function used to transform the data in *n*-dimensional space, the regularization penalty which specifies how much distance away from the hyperplane is allowed, and gamma, which determines how much influence points a given distance away have on the placement of the hyperplane. Support vector machines also do not require as much data as other models such as neural networks, are less likely to overfit the data, and are more likely to find the

Table 6
The full results of the models including the type of model, feature selection method used, dataset, MAPE, RMSE, and average computation time.

Model Type	Average Computation Time (sec)	Feature Selection Method	MAPE (Full, Microchannel, Unicategorical)	RMSE (Full, Microchannel, Unicategorical
Linear	3.5	PCA	46.70%, 43.40%, 44.90%	26851.1, 26354.6, 5149.6
		Boruta	26.35%, 28.10%, 23.80%	23146.0, 26317.2, 3688.0
		Recursive Partitioning	40.00%, 39.10%, 33.90%	23335.7, 25338.5, 4272.4
		Recursive Feature Elimination	29.9%, 35.50%, 828.80%	23261.5, 24970.3, 76620.4
		LASSO	26.95%, 27.06%, 22.73%	23207.3, 25107.8, 3630.7
		Stepwise Selection	27.68%, 28.44%, 22.40%	23080.0, 25076.6, 3549.2
GAM	20	PCA	36.18%, 35.72%, 40.57%	7803.9, 6924.3, 4568.0
		Boruta	18.32%, 21.60%, 17.29%	4071.7, 5182.2, 3158.4
		Recursive Partitioning	26.77%, 27.91%, 26.31%	4460.0, 4784.9, 3936.0
		Recursive Feature Elimination	20.99%, 26.11%, 29.03%	4395.7, 5690.2, 3980.0
		LASSO	20.48%, 18.82%, 16.89%	4120.5, 3845.1, 3040.6
		Stepwise Selection	20.56%, 21.28%, 18.02%	3781.7, 4117.4, 3002.4
MARS	195	PCA	35.01%, 5.67×10^8%, 44.75%	6690.7,8.4×10^9, 8972.0
		Boruta	20.47%, 21.41%, 14.36%	3852.7, 3896.7, 2455.7
		Recursive Partitioning	298.88%, 75.78%, 25.73%	26281.7, 9999.1, 3738.5
		Recursive Feature Elimination	22.45%, 27.97%, 33.74%	3944.1, 6563.4, 6942.8
		LASSO	19.86%, 19.92%, 14.34%	3367.3, 4244.2, 2396.6
		Stepwise Selection	20.03%, 20.06%, 16.19%	4994.2, 7873.4, 2295.5
Random Forest	252	PCA	35.04%, 34.13%, 32.53%	6445.7, 6744.0, 4754.7
		Boruta	16.14%, 16.23%, 15.25%	4473.6, 3409.4, 3395.9
		Recursive Partitioning	24.76%, 26.45%, 20.47%	5668.2, 6002.4, 3421.5
		Recursive Feature Elimination	13.51%, 17.84%, 14.37%	3107.2, 3348.5, 2408.2
		LASSO	15.30%, 13.97%, 13.05%	2726.0, 2766.7, 2458.6
		Stepwise Selection	15.32%, 15.50%, 11.94%	2854.2, 3226.8, 2060.4
GBM	407	PCA	40.39%, 44.58%, 41.24%	7605.9, 8924.7, 5217.8
		Boruta	19.60%, 22.63%, 15.16%	8175.0, 4425.3, 2126.3
		Recursive Partitioning	33.82%, 33.92%, 29.18%	7852.3, 8552.4, 4039.7
		Recursive Feature Elimination	22.14%, 21.49%, 17.70%	3506.2, 3849.9, 2388.4
		LASSO	14.03%, 14.27%, 17.50%	2829.7, 4576.6, 3365.0
		Stepwise Selection	20.46%, 20.01%, 12.98%	3534.4, 8321.2, 2153.0
SVM	562	PCA	56.40%, 51.18%, 46.97%	8689.1, 8938.5, 4695.1
		Boruta	11.32%, 16.13%, 14.98%	2542.8, 5039.1, 2864.7
		Recursive Partitioning	25.17%, 26.05%, 24.92%	5397.1, 5819.4, 3395.6
		Recursive Feature Elimination	17.78%, 22.43%, 16.82%	4932.8, 5355.3, 3053.5
		LASSO	16.12%, 11.02%, 10.64%	4737.5, 2641.0, 2293.4
		Stepwise Selection	17.15%, 16.26%, 14.80%	4816.2, 5019.3, 2882.7
Neural Network	878	PCA	70.83%, 72.45%, 47.70%	9890.1, 10284.8, 4937.1
		Boruta	13.57%, 14.94%, 14.54%	3248.6, 4332.0, 2270.1
		Recursive Partitioning	23.79%, 23.72%, 23.93%	4400.6, 4364.4, 3520.4
		Recursive Feature Elimination	15.20%, 17.54%, 13.34%	5853.9, 3549.7, 2536.0
		LASSO	16.15%, 16.39%, 13.73%	3847.5, 5170.9, 2378.4
		Stepwise Selection	15.88%, 14.23%, 14.66%	3696.6, 3835.2, 2645.1

global optimum in the dataset. For the full dataset, the Boruta selection method produced the best full dataset model with an MAPE of 11.32%. Very few observations, including those with large heat transfer coefficients, resulted in predicted value outside of a 30% prediction error. The model used a radial basis function as the kernel and approximately 7583 individual support vectors to predict the heat transfer coefficient. For the microchannel dataset, the LASSO selection method produced the best model with an MAPE of 11.02%. The model used a radial basis function as the kernel and approximately 7302 individual support vectors to predict the heat transfer coefficient. Overall, the microchannel dataset typically resulted in models that with an average MAPE 2–3% worse than the full dataset and had very similar patterns of overall Pearson correlations (Fig. 6) and feature importance plots (Fig. 10). It is unlikely that the microchannel dataset has significant usefulness when it comes to predicting untrained data, and further discussion and analysis will only address the full and unicategorical datasets as relevant. For the unicategorical dataset, the LASSO selection method produced the best overall model with an MAPE of 10.64%. This is the best model conducted with this fluid database. The model used a radial basis function as the kernel and approximately 2058 individual support vectors to predict the heat transfer coefficient. This suggests it was almost four times as easy for the support vector machine to map the unicategorical data as the full dataset. Some of this could be due to having less data to categorize while some can also be contributed to fewer outliers and uneven

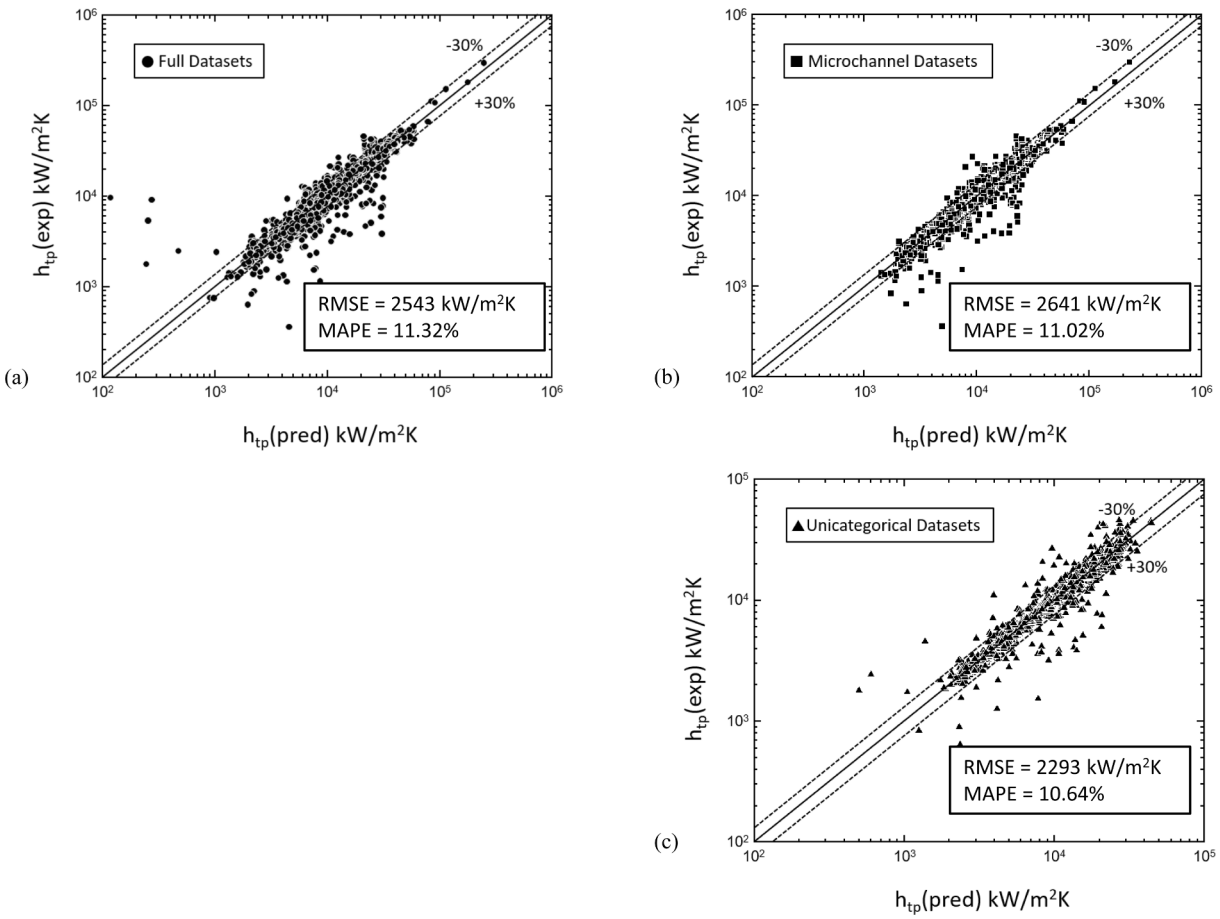

Fig. 9. Experimental vs. Predicted heat transfer coefficients for the best performing models (Support Vector Machines) displaying RMSE and MAPE for the (a) Boruta Selection Method - Full Dataset, (b) LASSO Selection Method - Microchannel Dataset, and (c) LASSO Selection Method - Unicategorical Dataset.

distributions making the data easier to model.

Looking at Table 6, we can see that a number of other models were able to achieve MAPE values only 2–3 percent higher than the support vector machines. The support vector machines use the Radial Basis Function Kernel which is a function that tries to measure the Euclidian distance between the input data in parametrized space and a fixed point representing the actual heat transfer coefficient. The only tuning parameters for support vector machines involve specifying the influence of a single data point on the ability to shift a support vector, and the penalty incurred in the model when an observation falls outside the determined maximum Euclidian distance. Both neural networks and random forest models have a significantly higher number of tuning parameters that require adjusting sample sizes, specifying the number of layers or decisions trees, applying regularization parameters at each layer, weighing specific features, etc. As a result, there is a higher possibility that the MAPE values reached using the random forest models and the neural networks are local minima instead of global minima.

In this study containing a moderately sized dataset with a large number of potential input variables, support vector machines performed well because they have a nonlinear mapping function, and they are tuned to find a balance between minimizing the effect of outliers while penalizing inaccuracy for each feature. For a medium-sized dataset with a large number of possible input variables the support vector machine was better at finding this balance than the neural network or random forest models, which can be more effected by unevenly distributed input variables, outlier values, and sampling sizes/methods that are defined by the user. Tuning models is a time-consuming process with no automated way of quickly guaranteeing the optimal set of parameters. It is possible that with additional time and a different approach, higher performing random forest models and neural networks could be achieved. Additionally, the large number of user-defined tuning parameters for neural networks make them significantly more computationally expensive. As seen in Table 6, the neural networks took an average of 56% longer to train and run than support vector machines.

Another important feature in determining the efficacy of these models is interpretability. The Bertsch and Kim & Mudawar correlations have an MAPE two to four times higher than the support vector machine or neural network, and they're limited to a specific range of fluids and experimental conditions, but they're extremely interpretable when it comes to understanding the mathematical relationships between experimental parameters and the heat transfer coefficient. They have a mathematical relationship with a known range of error. Support vector machines build a nonlinear mathematical relationship based on Euclidian distance in parametrized space, but it is difficult to determine exact numerical relationships for each variable. With a more specific kernel function and fewer input variables, it is possible to extract a mathematical relationship between the input variables and the heat transfer coefficient, but with such a high level of dimensionality, normalization, and transformation in this dataset, it's only feasible to approximate trends. Neural networks are even more difficult, as the combination of activation functions, multiple transformation layers, random sampling, and regularization makes it even more difficult to determine relationships between the input variables and the output. Machine learning can improve numerical accuracy at the expense of interpretability. Lowering the dimensionality of the input variables and adjusting the kernel function to reflect known mathematical relationships between the input variables and the heat transfer coefficient in the future may help bridge the gap between semi-empirical correlations and

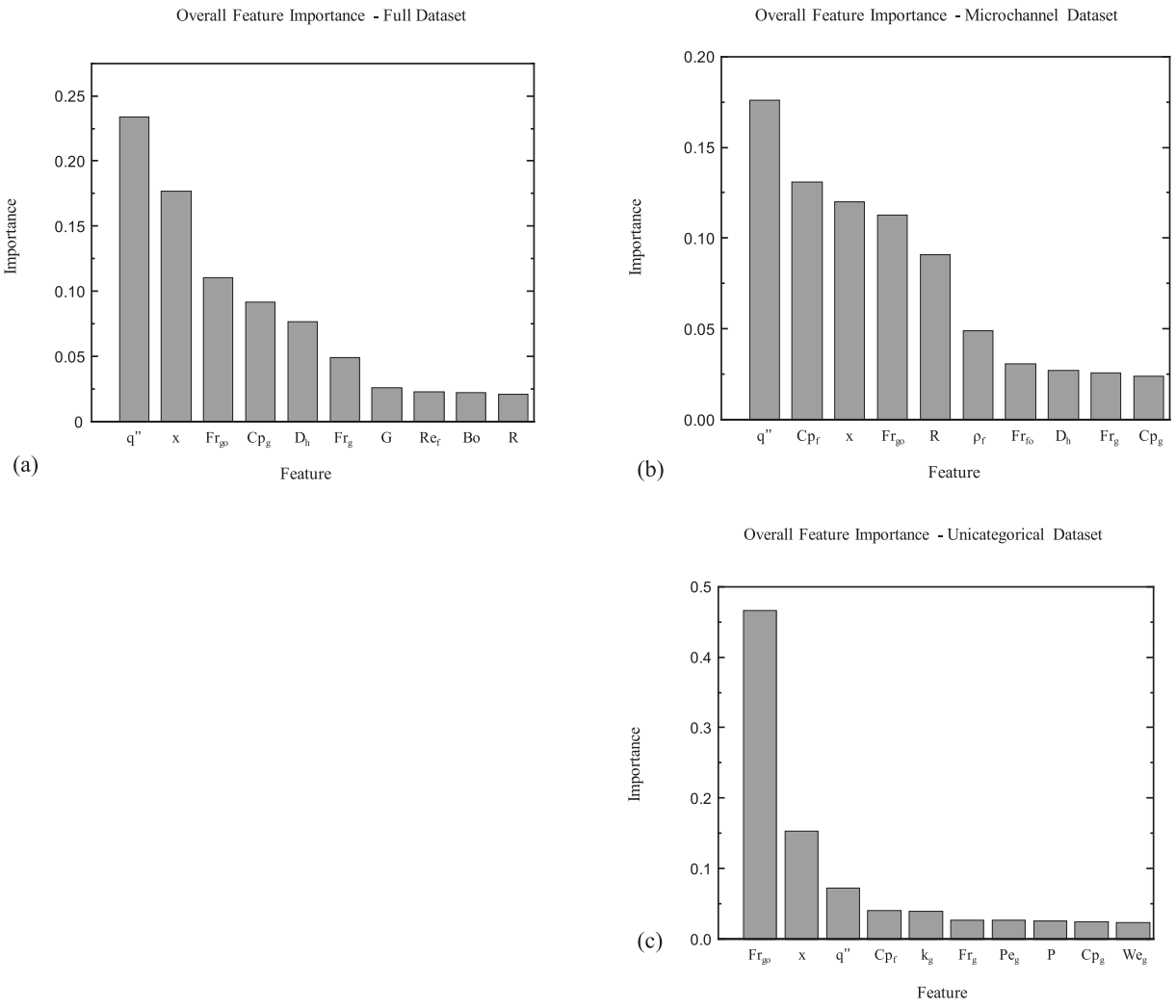

Fig. 10. Aggregated Feature Importance for the top 10 most significant variables across all models for the a) full, b) microchannel, and c) unicategorical datasets.

machine learning. Ultimately, the user must decide on the acceptable balance between interpretability and accuracy.

5.3. *Significant feature contributors*

Using the Pearson's correlation plots in Fig. 6 as an initial indicator and the aggregate feature importance plots in Fig. 10, it is possible to identify the variables with the greatest overall contribution to a large number of models. Looking at the average variable importance across all models, the quality, heat flux, and vapor-only Froude number were the most significant contributing features to the models in predicting the heat transfer coefficient. Further analysis of these top contributors can

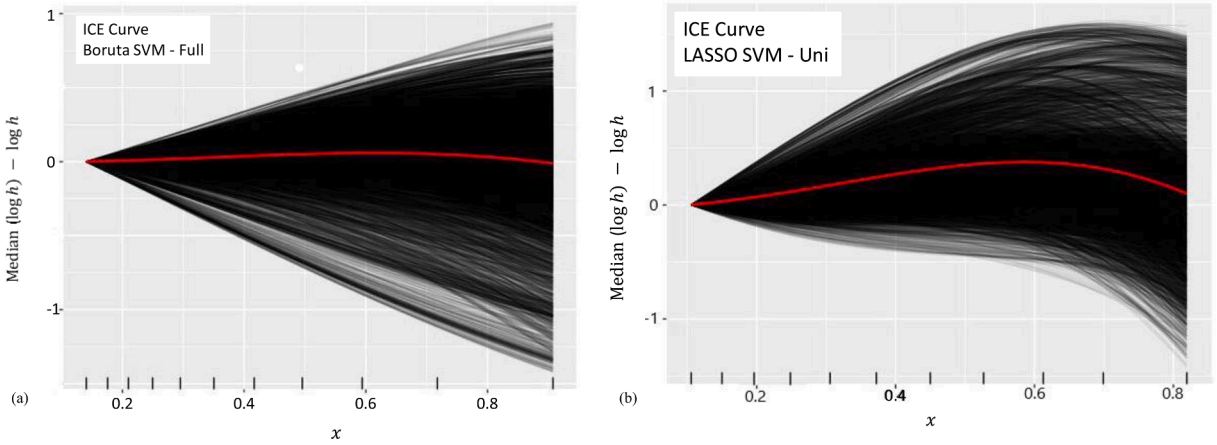

Fig. 11. Individual Conditional Expectation (ICE) curves for Quality () for a support vector machine with the (a) full dataset using the Boruta selection method and (b) Unicategorical dataset using the LASSO selection method.

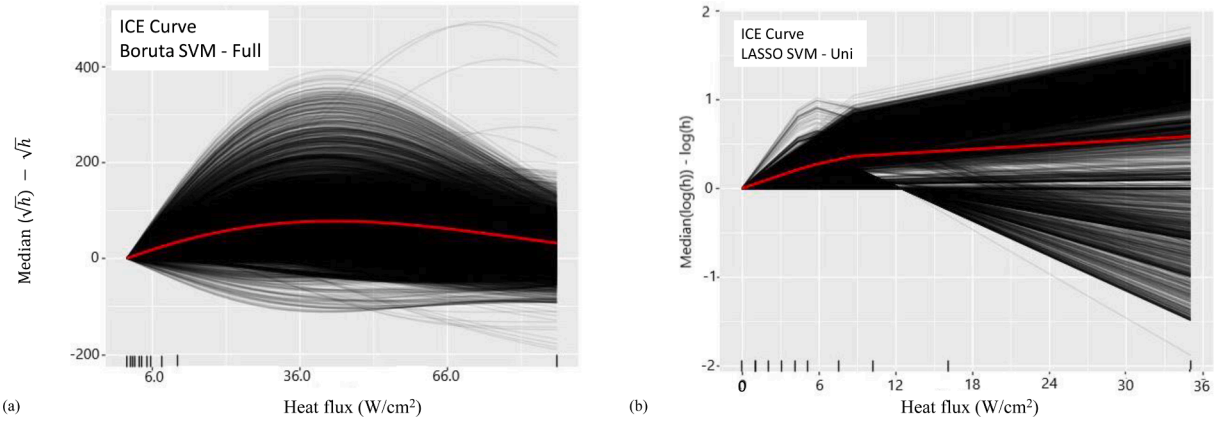

Fig. 12. Individual Conditional Expectation (ICE) curves for the heat flux for a support vector machine with the (a) full dataset using the Boruta selection method and (b) Unicategorical dataset using the LASSO selection method.

be done using Individual Conditional Expectation (ICE) curves, which show how the heat transfer coefficient is likely to change across the full range of values for a given individual variable. The depending variable is not the heat transfer coefficient directly, but rather a normalized value equivalent to the transformed dependent variable subtracted from median of the transformed dependent variable. This a more direct measurement of the influence a given feature has over a modeled variable when compared to the average (or when the given feature has no influence at all).

One of the top contributing features when predicting the heat transfer coefficient is the quality, indicating that the proportion of the fluid that is a saturated liquid is very influential in the fluid's effectiveness when transferring heat. A fluid can more effectively transfer heat within the two-phase regime, so the closer a fluid's quality is to a saturated liquid or saturated vapor, the lower the heat transfer coefficient. This can be seen in the ICE curves for the full and unicategorical dataset in Fig. 11(a) and (b). Higher qualities tend to cause a sharper decrease in the heat transfer coefficient, possibly do to the dryout phenomena where an increased relative amount of vapor can lead to less fluid on the edges of the channel which can lower the fluid's overall ability to transfer heat. Looking at the major inflection points for quality from the support vector machine models using the LASSO selection method for the unicategorical dataset, it appears that inflection points occur at qualities of approximately 0.2, 0.58, and 0.75 respectively. These qualities occur at a wide variety of experimental setup conditions and working fluids. The overall trend is that the heat transfer coefficient increases with quality until x is approximately 0.75, where there is a sharp decrease. This behavior seems to correspond to transition in flow boiling regimes discussed in [102] with bubbly flow, annular flow, and dryout point being the major influencers.

Another flow property not solely dependent on the experimental conditions or the working fluid is the heat flux. The impact of the heat flux can be seen in the ICE curves for the full and unicategorical dataset in Fig. 12(a) and 12(b). There are inflection points that appear to occur when the heat flux is approximately 6, 10, 14 W/cm^2 in the unicategorical dataset and 42.6, and 82.5 W/cm^2 for the full dataset. The two values in the full dataset are likely heavily driven by the Balasubramanian experiment [64], which reports heat flux values significantly higher than many of the other experiments and is not included in the unicategorical dataset. The lower three values occur at a wide variety of experimental setup conditions and working fluids. Two trends are clearly observed in the curves: first, part of the data shows a drop in heat transfer after around 5.9 W/cm^2 corresponding to the nucleate dominant boiling behavior, which is expected from dominant bubbly and slug flow regimes, and second, the remaining data shows an increase in heat transfer after 10 W/cm^2 corresponding to the convective dominant boiling behavior, which signifies a short bubbly and a long dominant annular flow regime [94]. In addition, we can see that the heat transfer coefficient has a strong positive relationship as the heat flux starts to increase, but it quickly levels off into an independent relationship.

The primary contributing flow characteristic is the Froude number, specifically the vapor-only Froude number, which describes the relative flow inertia to the external field. The impact of the vapor-only Froude number can be seen in the ICE curves for the full and unicategorical dataset in Fig. 13(a) and (b). There are inflection points appear to occur when the vapor-only Froude numbers of 2.5e7, 1.1e8, 1.5e8, 2.6e8, and

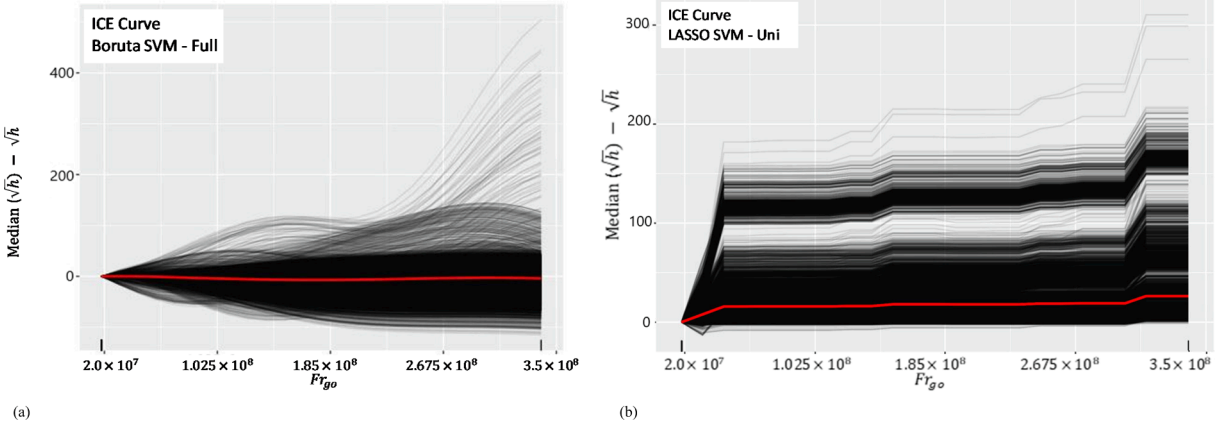

Fig. 13. Individual Conditional Expectation (ICE) curves for the vapor-only Froude number (Fr_{go}) for a support vector machine with the (a) full dataset using the Boruta selection method and (b) unicategorical dataset using the LASSO selection method.

3.3e8 across both models. Many of the values in the full dataset are heavily influenced by the outliers in the dataset and the Balasubramanian experiment [64]. Balasubramanian experiment contained significantly higher than average values for vapor-only Froude numbers across a large range of the temperature and fluid domain which resulted in a larger influence on the heat transfer coefficient. In the unicategorical dataset, however, the Balasubramanian experiment is not included, and the result is that the Froude number has a significant impact on the heat transfer coefficient at lower values as shown in the sharper initial increase at 2.5e7. Overall additional testing is necessary to determine the specific nature and numerical values related the trends between many of these features and the heat transfer coefficient and investigate the impact of individual experiments that weren't identified as outliers in the exploratory data analysis process such as the Balasubramanian experiment [64].

This study represents a thorough examination of feature engineering and numerical modeling techniques and their potential usefulness when predicting heat transfer coefficients in microchannels. Designing and executing physical experiments with microchannels and in thermal engineering at large is typically a costly and intensive process. There is great utility in machine learning and feature engineering when it comes to constricting the potential controlled variables and the boundary conditions of experimental properties. Many thermal engineering problems are complex analyses with nonlinear components and different effects late in time. Rather than relying on human factors such as previous experience and observations of past experiments, machine learning and feature engineering can be used to refine a smaller set of physical control variables such as channel diameter, channel length, type of fluid, temperature, and pressure. Even if the machine learning model cannot satisfactorily predict the heat transfer coefficient with enough accuracy to place confidence in the simulation alone, narrowing the physical design criteria is useful for many complex processes in thermal engineering.

6. Conclusions

This research was dedicated to developing better models for predicting the heat transfer coefficient for flow boiling in mini/microchannels using machine learning methods. The data used consisted of a fluid database of experimental data with 16,953 observations from 50 experiments with a total of 12 working fluids. Six different variable selection methods were used to obtain a varying number and distribution of features to include when modeling. Seven different modeling types were used to try and obtain the best predictions for the heat transfer coefficient. The key findings are as follows:

1. PCA and recursive partitioning did not seem to capture enough variance in the dataset, while the other four selection methods, including Boruta, recursive feature elimination, LASSO and stepwise selection, had mixed success depending on the model and subset of data being used.
2. The best overall model was obtained using a SVM with the LASSO selection method on the unicategorical dataset and achieved an MAPE value of 10.64%, while the best model trained on the full database used a support vector machine with the Boruta selection method on the full dataset and achieved an MAPE value of 11.33%.
3. The prediction capability of machine learning models, particularly SVM, appeared to far exceed the capabilities of semi-empirical correlations such as the Kim and Mudawar correlation.
4. The Froude number, the heat flux, and the quality are significant contributors to over 90% of all models performed using this fluids database. The mass velocity, channel roughness, and pressure coefficients $C_{p,f}$ and $C_{p,g}$ are also significant contributors in many models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This project was partly supported by the Office of Naval Research (ONR), under Grant Number N00014–21–1–2078, and by the National Science Foundation (NSF), under Grant Number 2138247. The views and conclusions contained herein are those of the authors only and should not be interpreted as representing those of ONR, the U.S. Navy or the U.S. Government.

Appendix A. Supplementary data

Feature selection summaries for each variable and feature selection method are provided in Table A.1. Supplementary data to this article can be found online at https://doi.org/10.1016/j.applthermaleng.2022.118305.

References

[1] L. Cheng, G. Ribatski, J.R. Thome, Two-Phase Flow Patterns and Flow-Pattern Maps: Fundamentals and Applications, Appl. Mech. Rev. 61 (2008), https://doi.org/10.1115/1.2955990.
[2] T.G. Karayiannis, M.M. Mahmoud, Flow boiling in microchannels: fundamentals and applications, Appl. Therm. Eng. 115 (2017) 1372–1397, https://doi.org/10.1016/j.applthermaleng.2016.08.063.
[3] J.R. Thome, L. Consolini, Mechanisms of Boiling in Microchannels: Critical Assessment, in: S. Kakaç, B. Kosoy, D. Li, A. Pramuanjaroenkij (Eds.), Microfluidics Based Microsystems, Springer, Netherlands, Dordrecht, 2010, pp. 83–105.
[4] C.L. Ong, J.R. Thome, Macro-to-microchannel transition in two-phase flow: Part 2 – Flow boiling heat transfer and critical heat flux, Exp. Therm. Fluid Sci. 35 (6) (2011) 873–886.
[5] S. Grauso, R. Mastrullo, A.W. Mauro, G.P. Vanoli, Flow boiling of R410A and CO2 from low to medium reduced pressures in macro channels: experiments and assessment of prediction methods, Int. J. Heat Mass Transf. 56 (1-2) (2013) 107–118.
[6] M.G. Cooper, Saturation nucleate pool boiling - a simple correlation, Inst. Chem. Eng. Symp. Ser. (1984) 785–793, https://doi.org/10.1016/B978-0-85295-175-0.50013-8.
[7] K.E. Gungor, R.H.S. Winterton, A general correlation for flow boiling in tubes and annuli, Int. J. Heat Mass Transf. 29 (3) (1986) 351–358, https://doi.org/10.1016/0017-9310(86)90205-X.
[8] Z. Liu, R.H.S. Winterton, A general correlation for saturated and subcooled flow boiling in tubes and annuli, based on a nucleate pool boiling equation, Int. J. Heat Mass Transf. 34 (11) (1991) 2759–2766, https://doi.org/10.1016/0017-9310(91)90234-6.
[9] S.M. Kim, I. Mudawar, Review of databases and predictive methods for heat transfer in condensing and boiling mini/micro-channel flows, Int. J. Heat Mass Transf. 77 (2014) 627–652, https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2014.05.036.
[10] Y. Liu, N. Dinh, Y. Sato, B. Niceno, Data-driven modeling for boiling heat transfer: Using deep neural networks and high-fidelity simulation results, Appl. Therm. Eng. 144 (2018) 305–320, https://doi.org/10.1016/j.applthermaleng.2018.08.041.
[11] D. Liu, S.V. Garimella, Flow Boiling Heat Transfer in Microchannels, J. Heat Transfer 129 (2006) 1321–1332, https://doi.org/10.1115/1.2754944.
[12] L. Cheng, G. Xia, Fundamental issues, mechanisms and models of flow boiling heat transfer in microscale channels, Int. J. Heat Mass Transf. 108 (2017) 97–127, https://doi.org/10.1016/j.ijheatmasstransfer.2016.12.003.
[13] J.R. Thome, V. Dupont, A.M. Jacobi, Heat transfer model for evaporation in microchannels. Part I: Presentation of the model, Int. J. Heat Mass Transfer 47 (14-16) (2004) 3375–3385, https://doi.org/10.1016/j.ijheatmasstransfer.2004.01.006.
[14] T. Cubaud, C.-M. Ho, Transport of bubbles in square microchannels, Phys. Fluids 16 (12) (2004) 4575–4585, https://doi.org/10.1063/1.1813871.
[15] V. Dupont, J.R. Thome, A.M. Jacobi, Heat transfer model for evaporation in microchannels. Part II: Comparison with the database, Int. J. Heat Mass Transfer 47 (14-16) (2004) 3387–3401, https://doi.org/10.1016/j.ijheatmasstransfer.2004.01.007.
[16] G. Xie, B. Sunden, Q. Wang, L. Tang, Performance predictions of laminar and turbulent heat transfer and fluid flow of heat exchangers having large tube-diameter and large tube-row by artificial neural networks, Int. J. Heat Mass Transf. 52 (11-12) (2009) 2484–2497, https://doi.org/10.1016/j.ijheatmasstransfer.2008.10.036.

[17] G.M. Lazarek, S.H. Black, Evaporative heat transfer, pressure drop and critical heat flux in a small vertical tube with R-113, Int. J. Heat Mass Transf. 25 (7) (1982) 945–960, https://doi.org/10.1016/0017-9310(82)90070-9.

[18] M.M. Shah, CHART CORRELATION FOR SATURATED BOILING HEAT TRANSFER: EQUATIONS AND FURTHER STUDY, in: ASHRAE Transactions, ASHRAE, 1982, pp. 185–196.

[19] S.-M. Kim, I. Mudawar, Theoretical model for annular flow condensation in rectangular micro-channels, Int. J. Heat Mass Transf. 55 (4) (2012) 958–970.

[20] H. Ganapathy, A. Shooshtari, K. Choo, S. Dessiatoun, M. Alshehhi, M. Ohadi, Volume of fluid-based numerical modeling of condensation heat transfer and fluid flow characteristics in microchannels, Int. J. Heat Mass Transf. 65 (2013) 62–72.

[21] S. Chen, Z. Yang, Y. Duan, Y. Chen, D.i. Wu, Simulation of condensation flow in a rectangular microchannel, Chem. Eng. Process. Process Intensif. 76 (2014) 60–69.

[22] J. Zhang, W. Li, W.J. Minkowycz, Numerical simulation of condensation for R410A at varying saturation temperatures in mini/micro tubes, Numerical Heat Transfer, Part A: Applications 69 (5) (2016) 464–478.

[23] Y. Qiu, H. Lee, C.R. Kharangate, Computational investigation of annular flow condensation in microgravity with two-phase inlet conditions, Int. Commun. Heat Mass Transfer 118 (2020), 104877. https://doi.org/10.1016/j.icheatmasstransfer.2020.104877.

[24] C.R. Kharangate, I. Mudawar, Review of computational studies on boiling and condensation, Int. J. Heat Mass Transf. 108 (2017) 1164–1196, https://doi.org/10.1016/j.ijheatmasstransfer.2016.12.065.

[25] Tlatelpa-Becerro, Rico-Martínez, Castro-Goméz, Calderón-Ramírez, Artificial Neural Networks (ANN) and Kalman Filter Algorithms to Predict Output Temperatures on a Heat Exchanger, 2018.

[26] Y. Qiu, D. Garg, L. Zhou, C.R. Kharangate, S.-M. Kim, I. Mudawar, An artificial neural network model to predict mini/micro-channels saturated flow boiling heat transfer coefficient based on universal consolidated data, Int. J. Heat Mass Transf. 149 (2020) 119211, https://doi.org/10.1016/j.ijheatmasstransfer.2019.119211.

[27] M.W. Wambsganss, D.M. France, J.A. Jendrzejczyk, T.N. Tran, Boiling heat transfer in a horizontal small-diameter tube, J. Heat Transfer 115 (1993) 963–972.

[28] T.N. Tran, Pressure drop and heat transfer study of two-phase flow in small channels, 1999.

[29] C.-C. Wang, C.-S. Chiang, J.-G. Yu, An experimental study of in-tube evaporation of R-22 inside a 6.5-mm smooth tube, Int. J. Heat Fluid Flow 19 (3) (1998) 259–269, https://doi.org/10.1016/S0142-727X(98)00006-X.

[30] Y.-Y. Yan, T.-F. Lin, Evaporation heat transfer and pressure drop of refrigerant R-134a in a small pipe, Int. J. Heat Mass Transf. 41 (24) (1998) 4183–4194, https://doi.org/10.1016/S0017-9310(98)00127-6.

[31] Z.Y. Bao, D.F. Fletcher, B.S. Haynes, Flow boiling heat transfer of Freon R11 and HCFC123 in narrow passages, Int. J. Heat Mass Transf. 43 (18) (2000) 3347–3358, https://doi.org/10.1016/S0017-9310(99)00379-8.

[32] W. Qu, I. Mudawar, Flow boiling heat transfer in two-phase micro-channel heat sinks—I. Experimental investigation and assessment of correlation methods, Int. J. Heat Mass Transf. 46 (15) (2003) 2755–2771, https://doi.org/10.1016/S0017-9310(03)00041-3.

[33] B. Sumith, F. Kaminaga, K. Matsumura, Saturated flow boiling of water in a vertical small diameter tube, Exp. Therm. Fluid Sci. 27 (7) (2003) 789–801, https://doi.org/10.1016/S0894-1777(02)00317-5.

[34] R. Yun, Y. Kim, M. Soo Kim, Y. Choi, Boiling heat transfer and dryout phenomenon of CO2 in a horizontal smooth tube, Int. J. Heat Mass Transf. 46 (13) (2003) 2353–2361, https://doi.org/10.1016/S0017-9310(02)00540-9.

[35] X. Huo, L. Chen, Y.S. Tian, T.G. Karayiannis, Flow boiling and flow regimes in small diameter tubes, Appl. Therm. Eng. 24 (8-9) (2004) 1225–1239.

[36] J. Lee, I. Mudawar, Two-phase flow in high-heat-flux micro-channel heat sink for refrigeration cooling applications: Part II—heat transfer characteristics, Int. J. Heat Mass Transf. 48 (5) (2005) 941–955.

[37] S. Saitoh, H. Daiguji, E. Hihara, Effect of tube diameter on boiling heat transfer of R-134a in horizontal small-diameter tubes, Int. J. Heat Mass Transf. 48 (23-24) (2005) 4973–4984.

[38] R. Yun, Y. Kim, M.S. Kim, Convective boiling heat transfer characteristics of CO2 in microchannels, Int. J. Heat Mass Transf. 48 (2) (2005) 235–242.

[39] R. Muwanga, I. Hassan, A Flow Boiling Heat Transfer Investigation of FC-72 in a Microtube Using Liquid Crystal Thermography, J. Heat Transfer 129 (2006) 977–987, https://doi.org/10.1115/1.2728905.

[40] X. Zhao, P.K. Bansal, Flow boiling heat transfer characteristics of CO2 at low temperatures, Int. J. Refrig. 30 (6) (2007) 937–945.

[41] B. Agostini, J.R. Thome, M. Fabbri, B. Michel, D. Calmi, U. Kloter, High heat flux flow boiling in silicon multi-microchannels – Part I: Heat transfer characteristics of refrigerant R236fa, Int. J. Heat Mass Transf. 51 (21-22) (2008) 5400–5414.

[42] L. Consolini, Convective boiling heat transfer in a single micro-channel, EPFL PP - Lausanne, n.d. https://doi.org/10.5075/epfl-thesis-4024.

[43] S.S. Bertsch, E.A. Groll, S.V. Garimella, Effects of heat flux, mass flux, vapor quality, and saturation temperature on flow boiling heat transfer in microchannels, Int. J. Multiph. Flow 35 (2) (2009) 142–154.

[44] S. In, S. Jeong, Flow boiling heat transfer characteristics of R123 and R134a in a micro-channel, Int. J. Multiph. Flow 35 (11) (2009) 987–1000.

[45] R. Mastrullo, A.W. Mauro, A. Rosato, G.P. Vanoli, Carbon dioxide local heat transfer coefficients during flow boiling in a horizontal circular smooth tube, Int. J. Heat Mass Transf. 52 (19-20) (2009) 4184–4194.

[46] H. Ohta, K. Inoue, M. Ando, K. Watanabe, Experimental Investigation on Observed Scattering in Heat Transfer Characteristics for Flow Boiling in a Small Diameter Tube, Heat Transfer Eng. 30 (1-2) (2009) 19–27, https://doi.org/10.1080/01457630802290080.

[47] L. Wang, M. Chen, M. Groll, Flow Boiling Heat Transfer Characteristics of R134a in a Horizontal Mini Tube, J. Chem. Eng. Data 54 (9) (2009) 2638–2645, https://doi.org/10.1021/je900140w.

[48] M. Ducoulombier, Ebullition convective du dioxyde de carbone-étude expérimentale en micro-canal, Univ. of Insa Lyon, Ecole Doctorale MEGA, 2010.

[49] M. Hamdar, A. Zoughaib, D. Clodic, Flow boiling heat transfer and pressure drop of pure HFC-152a in a horizontal mini-channel, Int. J. Refrig. 33 (3) (2010) 566–577.

[50] C. Martin Callizo, Flow boiling heat transfer in single vertical channels of small diameter, 2010.

[51] C.L. Ong, Macro-to-microchannel transition in two-phase flow and evaporation, EPFL, 2010.

[52] C.B. Tibiriçá, G. Ribatski, Flow boiling heat transfer of R134a and R245fa in a 2.3 mm tube, Int. J. Heat Mass Transf. 53 (2010) 2459–2468, https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2010.01.038.

[53] R. Ali, B. Palm, M.H. Maqbool, Flow Boiling Heat Transfer Characteristics of a Minichannel up to Dryout Condition, J. Heat Transfer 133 (2011), https://doi.org/10.1115/1.4003669.

[54] K.H. Bang, K.K. Kim, S.K. Lee, B.W. Lee, Pressure effect on flow boiling heat transfer of water in minichannels, Int. J. Therm. Sci. 50 (3) (2011) 280–286.

[55] J.B. Copetti, M.H. Macagnan, F. Zinani, N.L.F. Kunsler, Flow boiling heat transfer and pressure drop of R-134a in a mini tube: an experimental investigation, Exp. Therm. Fluid Sci. 35 (4) (2011) 636–644.

[56] M.M. Mahmoud, T.G. Karayiannis, D.B.R. Kenning, Surface effects in flow boiling of R134a in microtubes, Int. J. Heat Mass Transf. 54 (15-16) (2011) 3334–3346.

[57] H.-K. Oh, C.-H. Son, Flow boiling heat transfer and pressure drop characteristics of CO2 in horizontal tube of 4.57-mm inner diameter, Appl. Therm. Eng. 31 (2-3) (2011) 163–172.

[58] H.-K. Oh, C.-H. Son, Evaporation flow pattern and heat transfer of R-22 and R-134a in small diameter tubes, Heat Mass Transf. 47 (6) (2011) 703–717, https://doi.org/10.1007/s00231-011-0761-4.

[59] J. Wu, T. Koettig, C.h. Franke, D. Helmer, T. Eisel, F. Haug, J. Bremer, Investigation of heat transfer and pressure drop of CO2 two-phase flow in a horizontal minichannel, Int. J. Heat Mass Transf. 54 (9-10) (2011) 2154–2162.

[60] E. Costa-Patry, J. Olivier, J.R. Thome, Heat transfer charcacteristics in a copper micro-evaporator and flow pattern-based prediction method for flow boiling in microchannels, Frontiers in Heat and Mass Transfer (FHMT) 3 (1) (2012), https://doi.org/10.5098/hmt.v3.1.3002.

[61] T.G. Karayiannis, M.M. Mahmoud, D.B.R. Kenning, A study of discrepancies in flow boiling results in small to microdiameter metallic tubes, Exp. Therm Fluid Sci. 36 (2012) 126–142, https://doi.org/10.1016/J.EXPTHERMFLUSCI.2011.09.005.

[62] M. Li, C. Dang, E. Hihara, Flow boiling heat transfer of HFO1234yf and R32 refrigerant mixtures in a smooth horizontal tube: Part I. Experimental investigation, Int. J. Heat Mass Transf. 55 (13-14) (2012) 3437–3446.

[63] C.B. Tibiriçá, G. Ribatski, J.R. Thome, Flow boiling characteristics for R1234ze (E) in 1.0 and 2.2 mm circular channels, J. Heat Transfer 134 (2012) 20906.

[64] K. Balasubramanian, M. Jagirdar, P.S. Lee, C.J. Teo, S.K. Chou, Experimental investigation of flow boiling heat transfer and instabilities in straight microchannels, Int. J. Heat Mass Transf. 66 (2013) 655–671, https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2013.07.050.

[65] D. Del Col, S. Bortolin, D. Torresin, A. Cavallini, Flow boiling of R1234yf in a 1 mm diameter channel, Int. J. Refrig. 36 (2) (2013) 353–362.

[66] S. Grauso, R. Mastrullo, A.W. Mauro, J.R. Thome, G.P. Vanoli, Flow pattern map, heat transfer and pressure drops during evaporation of R-1234ze(E) and R134a in a horizontal, circular smooth tube: Experiments and assessment of predictive methods, Int. J. Refrig. 36 (2) (2013) 478–491.

[67] F. Vakili-Farahani, B. Agostini, J.R. Thome, Experimental study on flow boiling heat transfer of multiport tubes with R245fa and R1234ze(E), Int. J. Refrig. 36 (2) (2013) 335–352.

[68] R. Charnay, R. Revellin, J. Bonjour, Flow boiling characteristics of R-245fa in a minichannel at medium saturation temperatures, Exp. Therm. Fluid Sci. 59 (2014) 184–194, https://doi.org/10.1016/J.EXPTHERMFLUSCI.2014.01.011.

[69] S. Wang, M.Q. Gong, G.F. Chen, Z.H. Sun, J.F. Wu, Two-phase heat transfer and pressure drop of propane during saturated flow boiling inside a horizontal tube, Int. J. Refrig. 41 (2014) 200–209, https://doi.org/10.1016/J.IJREFRIG.2013.03.019.

[70] Z. Anwar, B. Palm, R. Khodabandeh, Flow boiling heat transfer, pressure drop and dryout characteristics of R1234yf: Experimental results and predictions, Exp. Therm. Fluid Sci. 66 (2015) 137–149, https://doi.org/10.1016/J.EXPTHERMFLUSCI.2015.03.021.

[71] R. Charnay, R. Revellin, J. Bonjour, Flow boiling heat transfer in minichannels at high saturation temperatures: Part I - Experimental investigation and analysis of the heat transfer mechanisms, Int. J. Heat Mass Transf. 87 (2015) 636–652, https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2015.03.081.

[72] B. Markal, O. Aydin, M. Avci, An experimental investigation of saturated flow boiling heat transfer and pressure drop in square microchannels, Int. J. Refrig. 65 (2016) 1–11, https://doi.org/10.1016/J.IJREFRIG.2015.12.013.

[73] Y. Xu, X. Fang, G. Li, D. Li, Y. Yuan, An experimental study of flow boiling heat transfer of R134a and evaluation of existing correlations, Int. J. Heat Mass Transf. 92 (2016) 1143–1157, https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2015.09.044.

[74] D.F. Sempértegui-Tapia, G. Ribatski, The effect of the cross-sectional geometry on saturated flow boiling heat transfer in horizontal micro-scale channels, Exp.

Therm. Fluid Sci. 89 (2017) 98–109, https://doi.org/10.1016/J.EXPTHERMFLUSCI.2017.08.001.
[75] D.F. Sempértegui-Tapia, G. Ribatski, Flow boiling heat transfer of R134a and low GWP refrigerants in a horizontal micro-scale channel, Int. J. Heat Mass Transf. 108 (2017) 2417–2432, https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2017.01.036.
[76] E.M. Fayyadh, M.M. Mahmoud, K. Sefiane, T.G. Karayiannis, Flow boiling heat transfer of R134a in multi microchannels, Int. J. Heat Mass Transf. 110 (2017) 422–436, https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2017.03.057.
[77] R. Peng, Exploratory data analysis with R, Lulu. com, 2012.
[78] F. Chollet, J.J. Allaire, Deep Learning with R, Manning Publications, 2018.
[79] S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate Imputation by Chained Equations in R, J. Stat. Softw. 1 (3) (2011) 2011. https://doi.org/10.18637/jss.v045.i03.
[80] D.M. Diez, C.D. Barr, etinkaya Rundel, M.,[OpenIntro Statistics:], OpenIntro, Inc., Sl. 3 (2015).
[81] I. Park, S.-M. Kim, I. Mudawar, Experimental measurement and modeling of downflow condensation in a circular tube, Int. J. Heat Mass Transf. 57 (2) (2013) 567–581.
[82] A. Jović, K. Brkić, N. Bogunović, A review of feature selection methods with applications, in: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) 2015 (2015) 1200–1205, https://doi.org/10.1109/MIPRO.2015.7160458.
[83] C.R. Rao, The Use and Interpretation of Principal Component Analysis in Applied Research, Sankhyā: The Indian Journal of Statistics, Series A (1961-2002). 26 (1964) 329–358.
[84] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (9) (2014) 2812–2831, https://doi.org/10.1039/C3AY41907J.
[85] M.B. Kursa, W.R. Rudnicki, Feature selection with the Boruta package, J. Stat. Softw. 36 (2010) 1–13.
[86] N. Bhalaji, K.B.S. Kumar, C. Selvaraj, Empirical study of feature selection methods over classification algorithms, Int. J. Intell. Syst. Technol. Appl. 17 (1/2) (2018) 98, https://doi.org/10.1504/IJISTA.2018.091590.
[87] V. Fortino, P. Kinaret, N. Fyhrquist, H. Alenius, D. Greco, A Robust and Accurate Method for Feature Selection and Prioritization from Multi-Class OMICs Data, PLOS ONE. 9 (2014) e107801.
[88] B. Efron, Missing Data, Imputation, and the Bootstrap, J. Am. Stat. Assoc. 89 (426) (1994) 463–475, https://doi.org/10.1080/01621459.1994.10476768.
[89] E.F. Cook, L. Goldman, Empiric comparison of multivariate analytic techniques: Advantages and disadvantages of recursive partitioning analysis, J. Chronic Diseases 37 (9-10) (1984) 721–731, https://doi.org/10.1016/0021-9681(84)90041-9.
[90] Y. Jiang, D. Cooley, M.F. Wehner, Principal Component Analysis for Extremes and Application to US Precipitation, J. Clim. 33 (15) (2020) 6441–6451.
[91] L. Cheng, L. Liu, Boiling and two-phase flow phenomena of refrigerant-based nanofluids: Fundamentals, applications and challenges, in: International Journal of Refrigeration, Elsevier, 2013, pp. 421–446. https://doi.org/10.1016/j.ijrefrig.2012.11.010.
[92] H. Wang, B.J. Lengerich, B. Aragam, E.P. Xing, O. Stegle, Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data, Bioinformatics 35 (7) (2019) 1181–1187, https://doi.org/10.1093/bioinformatics/bty750.
[93] S. Makridakis, Accuracy measures: theoretical and practical concerns, Int. J. Forecast. 9 (4) (1993) 527–529, https://doi.org/10.1016/0169-2070(93)90079-3.
[94] J.D. Jobson, in: Multiple Linear Regression BT - Applied Multivariate Data Analysis: Regression and Experimental Design, Springer, New York, New York, NY, 1991, pp. 219–398, https://doi.org/10.1007/978-1-4612-0955-3_4.
[95] S.N. Wood, Generalized additive models: an introduction with R, CRC Press, 2017.
[96] J.H. Friedman, Multivariate Adaptive Regression Splines, The Annals of Statistics 19 (1991) 1–67.
[97] J.H. Friedman, C.B. Roosen, An introduction to multivariate adaptive regression splines, Stat. Methods Med. Res. 4 (3) (1995) 197–217, https://doi.org/10.1177/096228029500400303.
[98] C.W. Chang, N.T. Dinh, Classification of machine learning frameworks for data-driven thermal fluid models, Int. J. Therm. Sci. 135 (2019) 559–579, https://doi.org/10.1016/j.ijthermalsci.2018.09.002.
[99] O. Maimon, L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook, Springer-Verlag, New York, 2005.
[100] G. Scalabrin, M. Condosta, P. Marchi, Modeling flow boiling heat transfer of pure fluids through artificial neural networks, Int. J. Therm. Sci. 45 (7) (2006) 643–663, https://doi.org/10.1016/j.ijthermalsci.2005.09.009.
[101] S.S. Bertsch, E.A. Groll, S.V. Garimella, A composite heat transfer correlation for saturated flow boiling in small channels, Int. J. Heat Mass Transf. 52 (7-8) (2009) 2110–2118, https://doi.org/10.1016/j.ijheatmasstransfer.2008.10.022.
[102] S.M. Kim, I. Mudawar, Universal approach to predicting saturated flow boiling heat transfer in mini/micro-channels – Part II. Two-phase heat transfer coefficient, Int. J. Heat Mass Transfer 64 (2013) 1239–1256, https://doi.org/10.1016/j.ijheatmasstransfer.2013.04.014.