Exploratory machine learning strategies for predicting thermal conductivity of materials from transient plane source measurement

Master's thesis in Computer science and engineering

BITNOORI LEE

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2023

Master's thesis 2023

Exploratory machine learning strategies for predicting thermal conductivity of materials from transient plane source measurement

BITNOORI LEE

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2023

Exploratory machine learning strategies for predicting thermal conductivity of materials from transient plane source measurement

BITNOORI LEE

Supervisor: Sebastianus Cornelis Jacobus Bruinsma, Ph.D, CSE
Advisor: Besira Mihiretie, Ph.D, Hot Disk AB
Examiner: Pedro Petersen Moura Trancoso, CSE

Master's Thesis 2023
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2023

Exploratory machine learning strategies for predicting thermal conductivity of materials from transient plane source measurement

BITNOORI LEE
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

This study introduces the application of machine learning to the Hot Disk Transient Plane Source (TPS) method, aimed at enhancing the precision and efficiency of thermal conductivity prediction.

Comprising two distinct parts, Part I, the research addresses the prediction of thermal conductivity in low-density/high-insulation materials. Part II is the thermal conductivity measurement under high-temperature conditions with noise. Four prediction algorithms were systematically applied and assessed for accuracy to predict thermal conductivities. Experimental data obtained through the TPS method served as the basis for machine learning training data, augmented with simulated data to make up for insufficient data.

The outcomes of this study provide a conclusive response to a critical research question: Can machine learning accurately predict thermal conductivity from transient curves? In Part I, machine learning consistently and accurately predicts thermal conductivity for low-density/high-insulation materials devoid of CL values, underscoring its complementary utility. In Part II, machine learning demonstrates its proficiency in accurately predicting thermal conductivity, even in noisy transient curves at extreme temperatures. However, challenges stemming from insufficient data issues and the absence of reference points introduce variability in accuracy.

Keywords: Machine Learning, Supervised Learning, Predictive Modeling, TPS methods, Thermal conductivity, FEM simulation.

Acknowledgments

I express profound gratitude to the individuals and organizations whose steadfast support and expertise were instrumental in completing this research.

I sincerely appreciate Hot Disk AB's invaluable contributions, generously sharing their extensive knowledge and 25 years' worth of data, which played an integral role in shaping this study. Heartfelt thanks are extended to CEO Mattias Gustavsson and the entire Hot Disk AB team for their unwavering support and indispensable contributions throughout this endeavor.

I am grateful to Besira Mihiretie, our esteemed company advisor, whose profound expertise and practical insights significantly enriched this research. His unwavering commitment to the project has been truly remarkable.

Additionally, I appreciate the valuable guidance the examiner, Pedro Petersen Moura Trancoso, and the supervisor, Sebastianus Cornelis Jacobus Bruinsma, provided in computer science. Their devoted time and expertise have been invaluable.

Last but not least, I would like to express sincere gratitude to my family and friends who gave persistent support from a distance in Korea during my two-year master's course. Their support was the driving force for successfully completing the master's program, and it was immeasurably great.

BITNOORI LEE, Gothenburg, 2023-08-29

Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

Thermal conductivity is a critical physical parameter that quantifies the ability of a material to conduct heat, which is crucial for applications such as heat exchange and thermal management. Depending on the application, there are various ways of characterizing thermal conductivity. Recently, the Hot Disk Transient Plane Source (TPS) method has emerged as a primary tool for measuring thermal properties [1], as evidenced by the number of published articles utilizing this method, more than 5,000 as of early 2023[2].

The TPS method is a proprietary and distinctive approach employed for thermal conductivity measurement. It has gained widespread adoption due to its versatility, brief measurement duration, and simplified experimental setup [3][4]. The method entails heating a thin material layer using an electric source and recording the resulting temperature increase as a transient curve illustrated in Figure 1.1 (right). This procedure necessitates solving the heat equation about the sensor's geometry and fitting the acquired data from the transient curve to a theoretically derived equation. During this fitting process, compensations must be made for instrument delay and the specific heat capacity of the sensing unit, often involving visual scrutiny of the residual plot in data analysis.

The thin material used is usually the nickel double spiral foil (referred to as the sensor) and covered by kapton or mica, Figure 1.1(left). It is used for heating and recording the change in temperature of the sample due to joule heating[1] as current runs through it. The time evolution of the sensor depicted in Figure 1.1(right) depends on the surrounding material and the thermal property of a material can be deduced from the recorded temperature profile if the sensor is placed inside the material[5]. One of the fundamental quantities measured using the TPS method is the thermal conductivity of the surrounding material, which is temperature-dependent. However, extreme temperature measurements are challenging for various reasons, including the mechanical and chemical changes in the sensor component, i.e., the nickel and the kapton/mica surrounding the nickel foil for mechanical support. This can cause non-ideal transient curves (noisy curves).

When applying the TPS method for low-density/high-insulation material character-

[1]the process of heating a material by passing an electric current through it.

ization, it is necessary to determine compensation coefficients (which determine the amount of power lost or unaccounted for) as it affects the accuracy of the measured thermal conductivity values. The contact thermal resistance (CL) value is specific to each material and sensor combination and can be determined experimentally. CL value is needed only for low-density/high-insulation material as a standard measurement does not need the CL value because it does not need to compensate for the power loss.

This research is motivated by constraints and challenges in the conventional TPS method described above. We consider the integration of machine learning techniques to offer promising solutions. Firstly, machine learning algorithms can predict a specific value without relying on hard-to-obtain parameters. Additionally, machine learning can potentially reduce noise in prediction tasks, improving accuracy and reliability. Consequently, machine learning can overcome the limitations and challenges of conventional methods.

This novel study explores the unexplored application of machine learning in conjunction with the TPS method for thermal conductivity. While machine learning has been utilized in the broader field [6][7], its specific integration with TPS is a new endeavor. The research aims to innovate their business and gain a competitive edge in a complex market by leveraging machine learning as a complementary measure to the existing TPS method. This study represents a pioneering effort to enhance accuracy and reliability in thermal conductivity analysis, staying at the forefront of the field and meeting growing market demands.

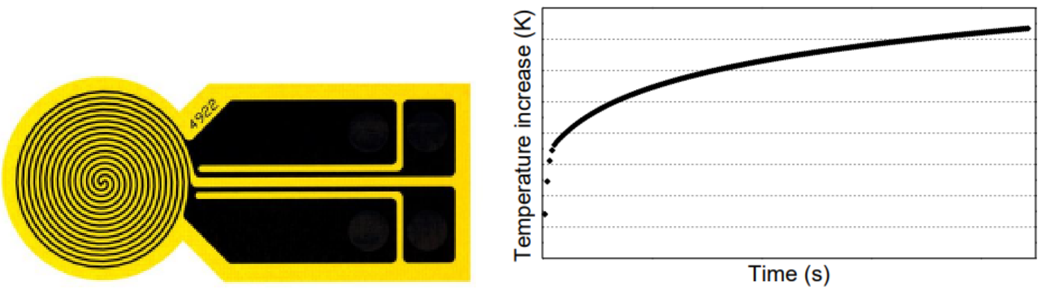

Figure 1.1: Hot Disk sensors with Kapton insulation (left) and A typical measurement transient (right)

1.2 Research Question

Can thermal conductivity be accurately predicted from transient curves using machine learning? The ultimate goal of the research is to reduce costs in thermal property measurement without sacrificing performance and to improve the speed and efficiency of the measurement process. Machine learning can potentially conserve resources, such as electrical energy for oven heating and machine operation, frequently wasted due to incorrect measurements and repeated trials.

The TPS method is now used for characterizing a wide range of materials, from very high conducting, such as copper, to extremely low conducting, such as Aerogel.

Also, it is used in a wide temperature range, from cryo (sub-zero) setup to 1000 °C.

In this study, two challenging applications of the TPS method were considered: low-density/high-insulation materials, which are very low conducting and requires unpractical compensation parameter for characterization. The second is the high-temperature region, which provides a non-ideal curve, resulting in the reproducible curve.

The research project is divided into two parts depending on the application considered:

Part I: Using the transient curve data of low-density/high-insulation material, thermal conductivity is predicted without CL value using different machine learning models such as regressions and Decision Tree. It is investigated whether thermal conductivity prediction through machine learning can demonstrate an accuracy equivalent to 5% standard deviation (STD) of the current thermal conductivity measurement accuracy using the Hot Disk method.

Part II: The goal is to improve the accuracy of predicting thermal conductivity by utilizing machine learning models in conjunction with noisy transient curve data obtained from high-temperature measurements and literature thermal conductivity values. The challenges associated with acquiring thermal conductivity values directly from the noisy transient curves obtained during measurements are intended to be overcome. Time-consuming iterations in the current method, necessitated by the particularly noisy high-temperature transient curves, will be addressed. Consequently, the feasibility of using machine learning techniques to enhance the precision of estimating thermal conductivity values at specific measurement temperatures will be explored, aiming for an accuracy better than the current 10% standard deviation (STD).

If actual measurement data is not accessible for both parts, simulated (synthetic) data is applied to achieve its objectives [8].

2

Theory

The theory section of this research paper provides an in-depth exploration of two fundamental aspects: the Hot Disk Theory and the application of machine learning techniques in the TPS method. Firstly, we delve into the Hot Disk Theory in 2.1, which encompasses the Hot Disk TPS method's advantages, working principle, and capability to measure thermal conductivity across various materials. We then transition into machine learning in 2.2, investigating the potential of leveraging regression techniques to predict thermal conductivity using transient curves. This section aims to provide a comprehensive understanding of both the Hot Disk Theory and the integration of machine learning in the TPS method, setting the stage for the subsequent analysis and evaluation presented in this study.

2.1 Hot Disk TPS Theory

The Hot Disk TPS method has many advantages over other thermal conductivity measurement techniques, including its ability to measure thermal conductivity and diffusivity of a wide range of materials, including liquids, solids, and powders, and its non-destructive and non-invasive nature. The technique is widely used in materials science, engineering, and other fields where thermal properties are essential.

The Hot Disk method is a technique for measuring materials' thermal conductivity and thermal diffusivity. The fundamental TPS technique uses a double spiral of conducting metal simultaneously as a continuous heat source and sensor. The spiral generates heat, which diffuses into the sample. Solving the heat conduction equation for the spiral geometry provides a relation between the change in the sensor temperature and the material's thermal conductivity. From this, the thermal conductivity and diffusivity of the material can be calculated using the following equations:

$$\overline{\Delta T}(\tau) = \frac{P_0}{\pi^{\frac{2}{3}} a K} D(\tau) \tag{2.1}$$

Where $\tau = \frac{\sqrt{\kappa t}}{a}$, K = $\kappa \rho c$, κ is thermal diffusivity, t is the test time, a is the radius of the largest ring in the sensor, P_0 is output power, K is thermal conductivity, ρ c is the volumetric specific heat of the material and $D(\tau)$ is a complex function of time[1], [9]. Observing the change in resistance during heating, the increase in sensor temperature is measured. The sensor's electrical resistance is correlated with the average temperature change across the sensor:

$$R = R_0[1 + \alpha\overline{\Delta T}(t)], \tag{2.2}$$

where R is the total resistance at time t, R_0 is initial resistance, α is the temperature coefficient of resistance of the spiral and $\overline{\Delta T}(t)$ is the average change in sensor temperature[1], [9]. Equation 2.1 shows that this increase in temperature depends on the thermal property of the surrounding material. The temperature evolution of the sensor versus time is fitted to Equation 2.1 to yield the thermal conductivity of the homogeneous material. However, a unique feature of the TPS method is that the thermal penetration depth can be determined as a function of the measurement time. This is described by the following relationships:

$$d_p = 2\sqrt{\kappa t} \tag{2.3}$$

where d_p represents the thermal depth of probing.

Thus, the mathematical model is expanded to approximate the thermal conductivity of inhomogeneous materials[10] using Equation 2.3. This is accomplished by taking into account shorter time intervals in the transient curve; hence, rather than fitting the transient curve entirely at once, the fitting is restricted to shorter time windows, $[t_i,t_{i+N}]$, where N is the total number of points in the newly reduced time window. The sample's thermal conductivity along the probing depth is evaluated by sliding the finite time window across the entire time range; this value is derived from the average estimated thermal diffusivity. An a priori known volumetric specific heat capacity of the sample solves the problem of fitting the model to the limited time window.

2.2 Machine Learning

Machine learning techniques have exhibited remarkable achievements in numerical value prediction [11],[12]. The objective of this study is to explore the prospective application of machine learning methods in the TPS method, which is divided into two distinct components: predicting thermal conductivity from transient curves without the need for CL values and predicting thermal conductivity from transient curves encompassing substantial noise encountered during measurements at extreme temperatures. This section briefly overviews several prediction algorithms employed in this study: linear regression, support vector regression (SVR), decision trees, and random forests. These techniques are utilized to assess the efficacy of predictive models.

Prediction is supervised learning that aims to predict a numerical value (dependent variable) based on the relationship between one or more independent variables. The main goal of the prediction is to develop a model that accurately predicts the dependent variable. The model is trained on training data, where the independent and dependent variables are known. Once the model is introduced, it can predict the dependent variable for new data. This study applied the following prediction algorithms to predict thermal conductivity: Linear Regression, Decision Tree, and Random Forest.

2.2.1 Linear Regression

Linear regression is a fundamental and widely used regression technique[13]. It assumes a linear relationship between the independent and dependent variables. The linear regression model fits a line through the data that minimizes the sum of squared errors between the predicted and actual values. This technique is proper under a linear relationship between the independent and dependent variables.

2.2.2 Support Vector Regression (SVR)

SVR is a prediction algorithm based on Support Vector Machines (SVMs) [14]. It is beneficial when the relationship between the independent and dependent variables is non-linear. The SVR model maps the data to a higher dimensional space where a linear relationship can be established. In situations with a non-linear association between independent and dependent variables, the model identifies a hyperplane that optimizes the gap between the predicted and actual values. This approach is valuable as it maximizes the margin in such cases.

2.2.3 Decision Tree

Decision Tree is a non-parametric prediction algorithm that creates a tree-like model of decisions and their potential consequences [15]. The model partitions the data into smaller subsets based on the values of the independent variables. The tree is built recursively by splitting the subsets until a stopping criterion is met. This technique is proper when the relationship between the independent and dependent variables is non-linear and complex.

2.2.4 Random Forest

Random Forest is an ensemble learning technique that combines multiple decision trees to improve the accuracy of the predictions [16]. It is useful when the relationship between the independent and dependent variables is non-linear and complex. The model creates multiple decision trees by randomly selecting subsets of the training data and independent variables. When confronted with a non-linear and intricate connection between independent and dependent variables, this method proves valuable as it involves computing the average of predictions from multiple decision trees to arrive at the final prediction.

2.3 Conclusion

To summarize this section, the Hot Disk TPS method is valuable for measuring thermal conductivity and diffusivity. Its non-destructive and non-invasive nature offers significant advantages in assessing a wide range of materials. Furthermore, by incorporating machine learning techniques, the TPS method can enhance its predictive capabilities, overcoming challenges such as the need for calibration values and noise in transient curves. Regression techniques provide a promising avenue for developing accurate and efficient models for estimating thermal conductivity. This integration of machine learning and the TPS method holds immense potential for advancing our understanding and characterization of thermal properties, benefiting various scientific and engineering disciplines.

3

Methods

The method section of this research paper covers various essential steps taken to conduct the study effectively. In the following subsections, an overview is provided for each step. Firstly, in 3.1, the process of collecting and preparing the data for analysis is outlined. This includes the collection of experimental data, simulated data, and the utilization of COMSOL Multiphysics simulation. Subsequently, in the 3.1.2, the details of how the collected data was processed and prepared for further analysis are delved into. This subsection is divided into two parts, focusing on preparing low-density/high-insulation materials and high-temperature measurements. Following this, in the 3.2, the criteria and methodology used to select and train the predictive models are discussed. Lastly, in the 3.3, the process of evaluating the trained models and optimizing their performance is outlined. This comprehensive method section provides the necessary details to ensure transparency and reproducibility in the study's methodology. It serves as a foundation for this research's subsequent analysis and findings.

3.1 Data Collection and Preparation

3.1.1 Data Collection

3.1.1.1 Experimental Data

The first step was to collect and prepare the data from accumulated experimental data that will be used for training the regression model. This involved cleaning the data and ensuring it was formatted so the regression algorithm could easily use it. The next step was to select the features or variables that would be used to train the model. These features should be relevant to the predictive task and strongly correlated with the target variable.

Data measured through the TPS method and Hot Disk software is provided as an Excel sheet containing two data categories.

Independent variables; Time evolution of temperature [t,T], Power(P), sensor resistance(R), Temperature coefficient of resistance(TCR)

Dependent variables; Thermal conductivity

In this study, thermal conductivity is the target value, and the related independent values were selected as the input features. A detailed explanation can be found in Section 3.1.2.

3.1.1.2 Simulated data

The data used in this study consist of measurements made using the TPS method over several decades. However, if the application is limited to a specific field, such as low-density/high-insulation samples, the utilization of naturally existing materials is limited. Also, the database of these materials needs to be bigger for machine learning methods to be used. To compensate for this, a machine learning application used synthetic data based on a detailed study [8] of a previous numerical simulation of the Hot Disk method. This was done by a collaborator, an industry doctoral student working under the same supervisor, using COMSOL multiphysics simulation software [17], which is used for solving complex problems in various fields of science and engineering. Additionally, numerous studies have provided documentation of the validity of COMSOL's heat transfer modules.[18]–[20]

3.1.1.3 COMSOL multiphysics simulation

Several steps were involved to simulate heat transfer problems and obtain simulated COMSOL data. Firstly, the model's geometry was defined, including importing a CAD file or creating a 3D model from scratch within the software, specifying dimensions, shapes, and boundaries needed to simulate the heat transfer problem accurately. Secondly, heat transfer physics was defined, including setting the governing equations and boundary conditions that describe how heat is transferred in the system, using various heat transfer models such as conduction, convection, and radiation. Thirdly, the model was meshed, which involves dividing the geometry into more minor, finite elements to discretize the problem and solve it numerically, using various meshing algorithms and options. Finally, the model was solved using COMSOL's built-in solvers, which use the finite element method to solve the partial differential equations that describe the heat transfer problem. The simulated data was visualized and analyzed using COMSOL's post-processing tools. Figure 3.1,3.2, shows visualized the results such as temperature distributions, heat fluxes, and temperature gradients.

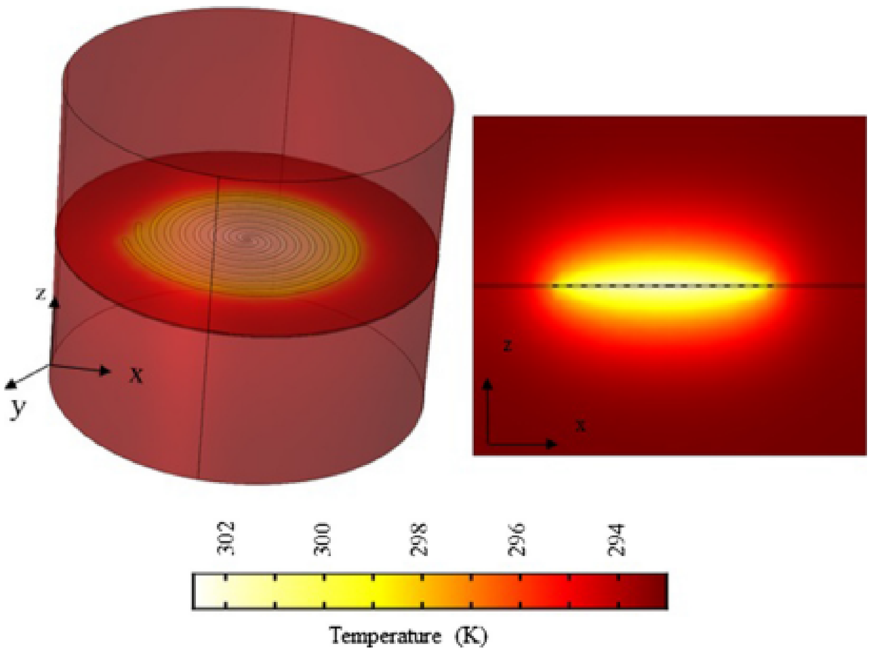

Figure 3.1: 3D temperature profile of the sensor and sample(left), and 2D view of 3D simulation of the sensor and sample(right) [8].

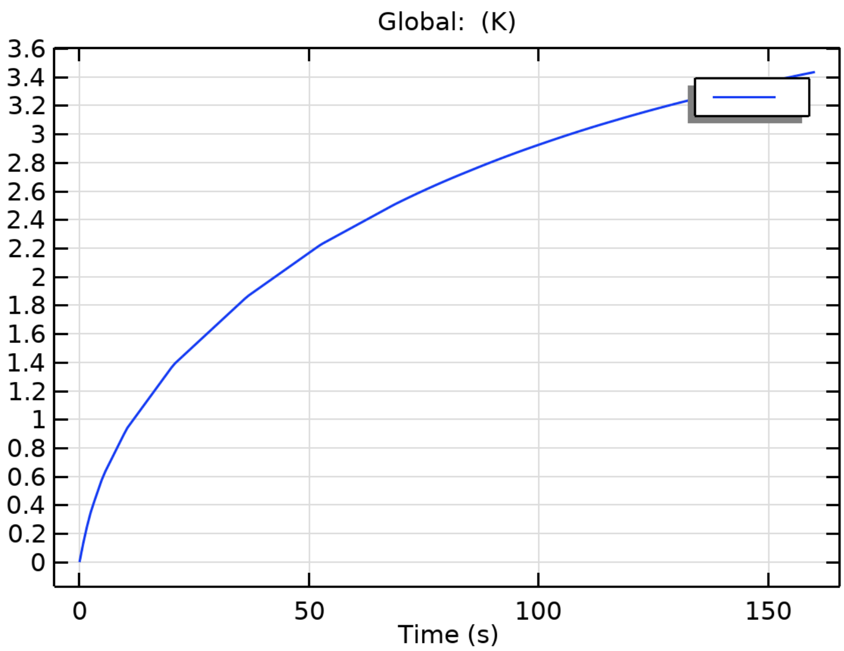

Figure 3.2: Example plot of temperature increase of sensor from the COMSOL simulation report.

3.1.2 Data Preparation

3.1.2.1 Part I: Low-density/high-insulation material

The prediction result in machine learning can potentially be affected by the order of selected data set features. This is because different machine learning algorithms treat the input features differently and may be sensitive to the ordering of features or columns.

For example, linear regression weights for each feature are based on the correlations between the elements and the target variable. Changing the ordering of features can result in different correlations being learned, affecting the prediction results. On the other hand, decision trees and random forests are not generally sensitive to the ordering of features, as they make their decisions based on the information gained of each feature independently of other features.

To minimize the impact of the order of arranging columns on the data set and to better understand how each algorithm treats the input features, four different orders of data sets consisting of 48 samples are prepared for Part I, which is described in Section 1.2 Low-density/high-insulation material:

Table 3.1: Datasets with input vectors in different orders

$t-T$	t_0	t_1	...	t_{100}	T_0	T_1	...	T_{100}	R_d	$Power_0$	r	TCR	λ
$\sqrt{t}-T$	$\sqrt{t_0}$	$\sqrt{t_1}$	...	$\sqrt{t_{100}}$	T_0	T_1	...	T_{100}	R_d	$Power_0$	r	TCR	λ
Shuffled	t_0	T_0	t_1	T_1	...	...	t_{100}	T_{100}	R_d	$Power_0$	r	TCR	λ
normalized	t_0	t_1	...	t_{100}	T'_0	T'_1	...	T'_{100}	R_d			TCR	λ

Where R_d is the resistance of the disk, $Power_0$ is the output power, r is the sensor radius, λ is the thermal conductivity and

$$T' = T * Power_0/r \tag{3.1}$$

3.1.2.2 Part II: High-temperature measurements

Figure 3.3 shows in-house measurements of the thermal conductivity of fused quartz characterized using the Hot Disk TPS method between -10 °C and about 900 °C, along with the literature values from 0 to 650 °C [21]

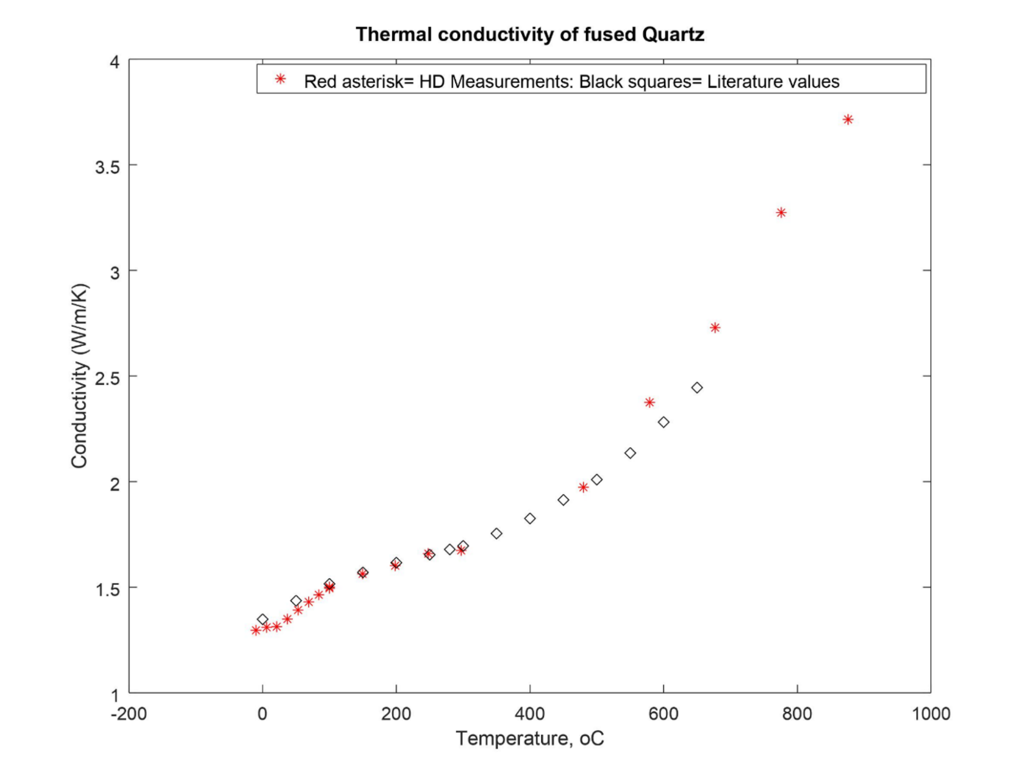

Figure 3.3: Thermal Conductivity of fused Quartz at different experimental temperatures

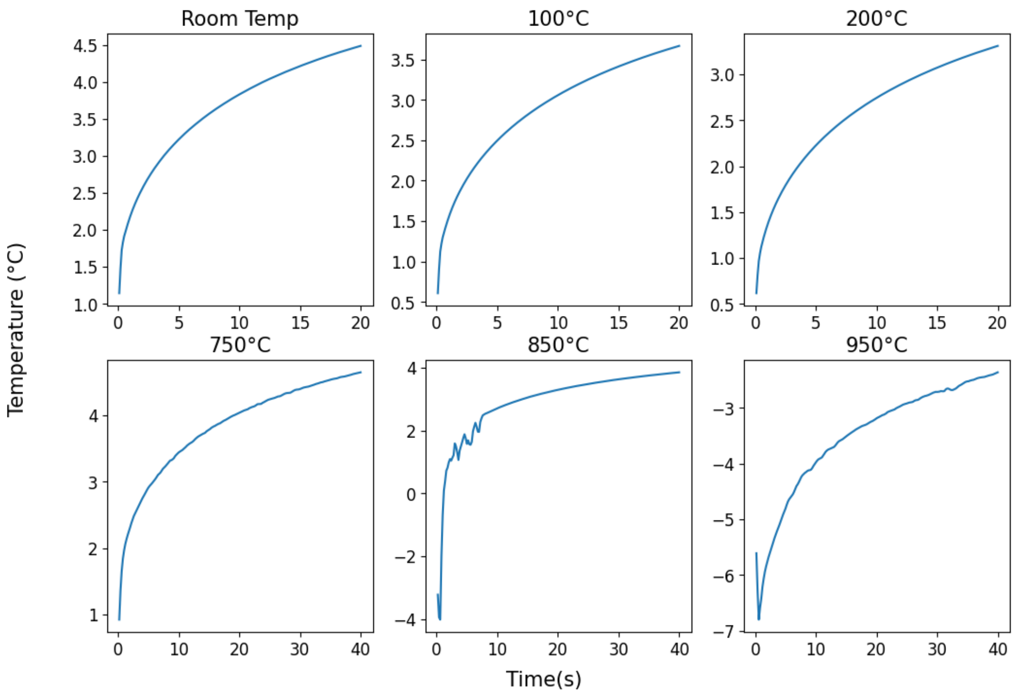

Figure 3.4: Noise variation of the transient curve of fused quartz at different measurement temperatures

From the curves over a temperature range of -10 to 1000 °C, it can be observed that a nice curve is exhibited by the measurements up to 650 °C, indicating that the measured thermal conductivity values can be trusted. However, the higher temperature range data exhibit varying noise levels, as shown in Figure 3.4.

Unlike Part I described in Section 1.2, it is interesting to note that when more than one measurement is included for noisy data at the same temperature and conditions, different levels of noise sometimes result from repeated measures.

The approach and data vectors employed in Part I were also utilized in Part II to predict thermal conductivity. However, a different methodology was applied for Part II, explicitly focusing on Quartz as a single sample. Experimental data obtained through the TPS method at various measured temperatures were used in this analysis. A polynomial fit was employed to compare the obtained results with the thermal conductivity values reported in the literature. The detailed results of this analysis can be found in Section 4.2.1.

By leveraging the TPS method and utilizing experimental data from Quartz at different temperatures, the goal was to derive insights into the thermal conductivity properties. The polynomial fit allowed for comparing the obtained results with existing literature values. For a comprehensive understanding of the findings, the following section describes the results obtained from this analysis.

3.2 Model Selection and Training

After selecting the features, the next step involved choosing the regression algorithms used to train the model. Several regression algorithms were chosen, including linear regression, SVR, decision trees, and random forest, an ensemble of unpruned regression trees. After each algorithm was selected, the model was trained on the prepared data. The model parameters can be adjusted to minimize the difference between the predicted values and the actual values in the training set during the following steps:

1. Import required libraries from pandas and sklearn

2. Load and Split the data into training and testing sets using train_test_split, where the testing set is 20% of the data, and the random state is set to 0 for reproducibility.

3. Fit a linear regression model to the training data for each of the 4 data sets mentioned in Section 3.1.2: t-T, sqrt t-T, shuffled, and normalized.

4. Use the fitted models to predict the target variable for the testing data for each of the four data sets.

3.3 Model evaluation and optimization

After each model was trained, the next step involved evaluating its performance. In this study, Mean squared error (MSE) and R-squared score were used to evaluate the performance of a machine learning model for prediction in regression problems. By using MSE and R-squared scores, a better understanding of the model's performance can be obtained. Mean squared error (MSE) is a common evaluation metric employed in machine learning to gauge the quality of a model's fit to the data. It computes the squared disparities' mean between the predicted and actual values.[22].

MSE can be calculated by taking the sum of the squared differences between the predicted and actual values and then dividing it by the number of data points. Mathematically, MSE is defined as:

$$MSE = 1/n * \sum (y_i - \hat{y}_i)^2 \tag{3.2}$$

where n is the number of data points, y_i is the actual value and $\hat{y}_i$ is the predicted value.

MSE is a non-negative value; a lower value indicates a better fit between the model and the data. It is a popular metric because it gives more weight to more significant errors and penalizes them more heavily. This means it is more sensitive to significant deviations between predicted and actual values than minor deviations.

To evaluate a machine learning model using MSE, the model is first trained on a training dataset and then tested on a separate testing dataset. The predicted values for the testing dataset are compared to the actual values, and the MSE is calculated.

The R-squared score, alternatively referred to as the coefficient of determination, quantifies the fraction of the variability in the dependent variable that can be accounted for by the independent variable within a regression model. [22]. It measures the goodness of fit of a linear regression model and how well it fits the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{3.3}$$

where y_i represents the actual values, $\hat{y}_i$ represents the predicted values, $\bar{y}$ represents the mean of the actual values, and n represents the number of observations.

The R-squared score ranges from 0 to 1, where 0 means the model does not explain any of the variance in the dependent variable, and 1 represents the model that explains all the variance. In other words, the closer the R-squared score is to 1, the better the model fits the data. To calculate the R-squared score, the sum of the squared differences between the predicted and actual values is divided by the sum of the squared differences between the actual values and the mean of the dependent variable. The resulting value is subtracted from 1 to get the R-squared score.

Generally, a higher R-squared score indicates a better fit of the model to the data. However, it is essential to note that the R-squared score may only sometimes be the best metric to evaluate the performance of a model, as it does not consider the complexity of the model and the potential overfitting of the data. Therefore, it should be combined with other metrics and techniques for model evaluation.

If the model's performance is unsatisfactory, it can be optimized by tweaking the algorithm's hyperparameters or adjusting the training features.

3.4 Conclusion

In conclusion, the method section of this research paper has provided a detailed account of the steps undertaken to conduct the study. The data collection process involved gathering experimental data, simulated data, and utilizing COMSOL multiphysics simulation. Following that, the collected data was prepared for analysis, considering both low-density/high-insulation materials and high-temperature measurements. Predictive model selection and training were conducted, adhering to specific criteria and methodologies. Finally, the trained models were evaluated and optimized for their performance and reliability. The method section serves as a comprehensive guide, ensuring transparency and reproducibility in the study's methodology and setting the stage for the subsequent analysis and findings presented in this research.

4

Results

The results section of this research paper presents the findings and outcomes of the study, with a focus on two main parts. In 4.1, the prediction results for low-density/high-insulation materials are examined. This includes an analysis of the predictions using only the experimental data, where the models are developed based on the collected experimental data. Additionally, the prediction results obtained by incorporating both experimental and simulated data are explored, aiming to enhance the reliability and accuracy of the predictive models. Moving on to 4.2, the focus shifts to high-temperature measurements, where the prediction results obtained using experimental data across various measurement temperatures are presented.

4.1 Part I: Low-density/high-insulation material

4.1.1 Prediction with experimental data

Linear regression, SVR, Decision Tree and Random Forest train the four datasets with 48 samples. Most results show that MSE is high, with a negative R-squared score, and the best R-squared score is 0.3, which is way below the range that can be considered accurate. From the initial result, it is found that most outliers are in high-conducting materials, and the distribution of conductivity values is analyzed as shown in Figure4.1. The data set is grouped by thermal conductivity into three samples: 0-50W$m^{-}1K^{-}1$, 0-20W$m^{-}1K^{-}1$, and 0-1W$m^{-}1K^{-}1$, with 39, 32, and 20 samples, respectively, and trained iteratively.

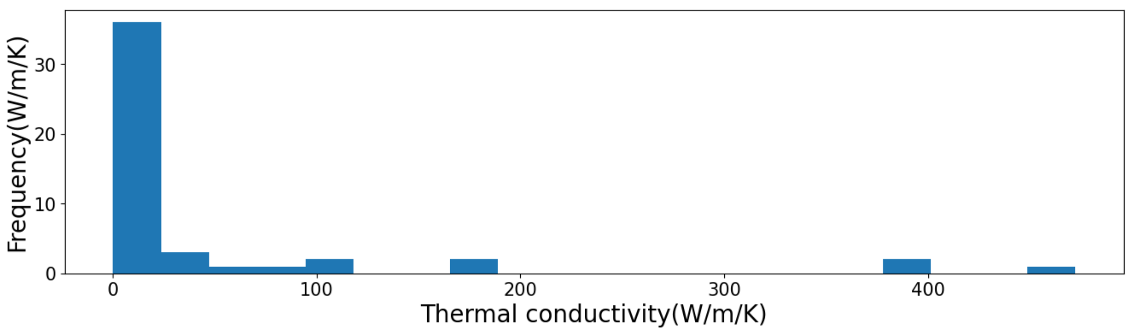

Figure 4.1: Distribution of the number of samples according to thermal conductivity

Figure 4.2 shows that the actual and predicted thermal conductivity values do not match. It can be seen that linear regression is unsuitable for predicting the sample data's thermal conductivity value.

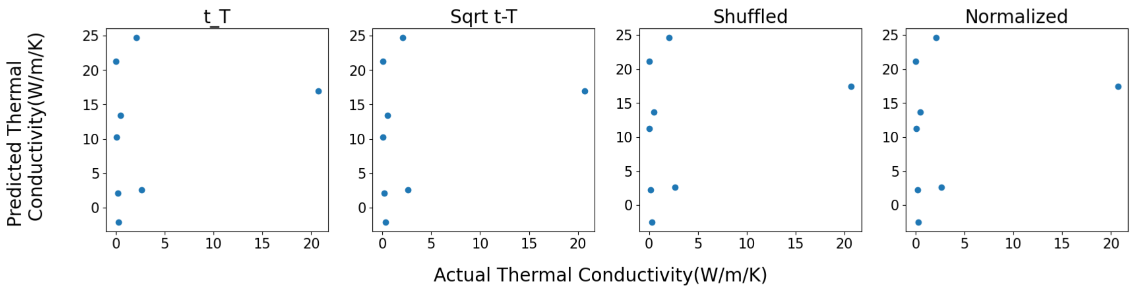

Figure 4.2: Result of training with Linear regression with data of 39 samples with a thermal conductivity lower than 50 W$m^{-}1K^{-}1$.

As shown in Figure 4.3, the actual and predicted thermal conductivity values are in excellent agreement regardless of the order of the column vectors. Negative R-squared scores obtained during the model training with Linear Regression and SVM indicate suboptimal performance due to weak feature correlation or the failure to capture non-linear relationships. One possibility is that the chosen features used for training the models are not sufficiently correlated with the target variable, leading to weak predictive performance. Another possibility is that the relationship between the features and the target variable is non-linear, while Linear Regression assumes a linear relationship. Consequently, the regression algorithm might encounter difficulties in capturing the intricacies of the data, potentially leading to poor performance. It can be speculated that the latter is more likely as the Decision tree shows the highest accuracy among applied algorithms for samples with a thermal conductivity lower than 50 W$m^{-}1K^{-}1$ as shown in Table 4.1.

To train different models removing outliers, such as samples with relatively high thermal conductivities, the data was sampled with a thermal conductivity of $20m^{-}1K^{-}1$ or less and proceeded like above. This investigation attempted to determine whether high prediction accuracy can be demonstrated in models other than the Decision Tree.

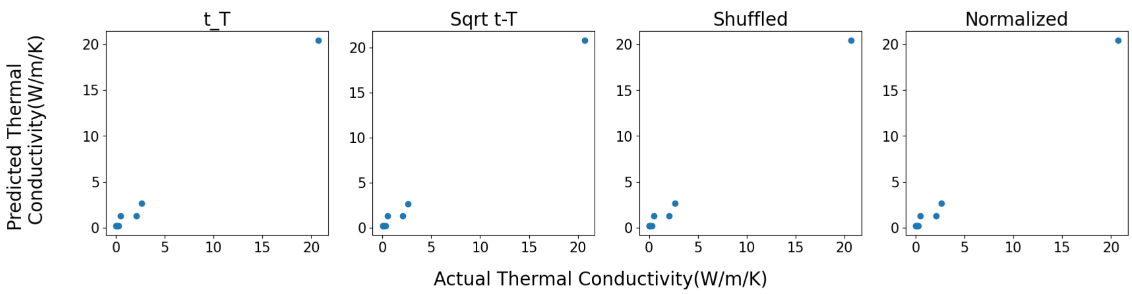

Figure 4.3: Result of training with Decision Tree with data of 39 samples with a thermal conductivity lower than 50 W$m^{-}1K^{-}1$.

Table 4.1: Evaluation results by dataset and regression algorithm of samples with thermal conductivity lower than 50 W$m^{-}1K^{-}1$

	MSE	**R-squared score**
Linear Regression		
t-T	157.004	−2.568
$\sqrt{t}$**-T**	157.004	−2.568
Shuffled	159.314	−2.62
Normalized	159.314	−2.62
SVR		
t-T	50.814	−0.155
$\sqrt{t}$**-T**	50.814	−0.155
Shuffled	50.814	−0.155
Normalized	50.814	−0.155
Decision Tree		
t-T	0.175	0.996
$\sqrt{t}$**-T**	0.176	0.996
Shuffled	0.175	0.996
Normalized	0.181	0.996
Random Forest		
t-T	0.424	0.99
$\sqrt{t}$**-T**	0.424	0.99
Shuffled	0.424	0.99
Normalized	0.424	0.99

Figure 4.4 shows the modeling results with a Random Forest of groups filtering out only samples with a thermal conductivity lower than $20m^-1K^-1$, which has a higher accuracy trend over $50m^-1$. Except for the Random forest using normalized data, the R-squared score is negative, indicating that the selected model does not follow the data trend as shown in Table 4.2.

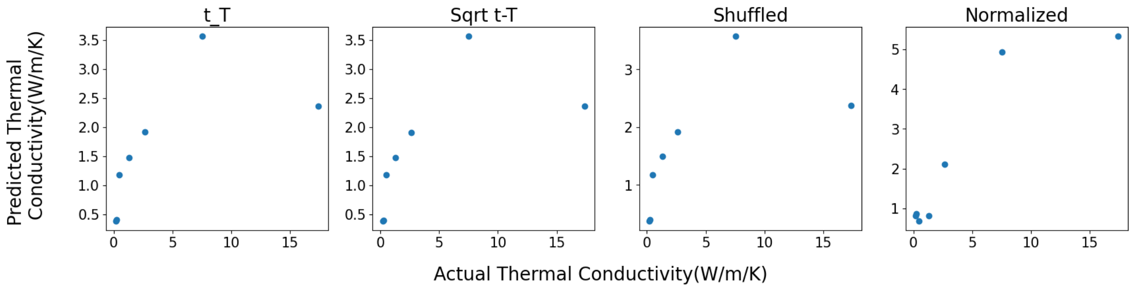

Figure 4.4: Result of training with Random Forest with data of 32 samples with a thermal conductivity lower than 20 $\mathrm{W}m^-1K^-1$.

Table 4.2: Evaluation results by dataset and regression algorithm of samples with thermal conductivity lower than 20 $\mathrm{W}m^-1K^-1$

	MSE	R-squared score
Linear Regression		
t-T	39.61	−0.153
$\sqrt{t}$**-T**	38.442	−0.119
Shuffled	45.05	−0.312
Normalized	23.719	0.309
SVR		
t-T	45.734	−0.332
$\sqrt{t}$**-T**	47.644	−0.387
Shuffled	45.734	−0.332
Normalized	45.427	−0.323
Decision Tree		
t-T	49.113	−0.43
$\sqrt{t}$**-T**	46.509	−0.354
Shuffled	46.218	−0.346
Normalized	47.056	−0.37
Random Forest		
t-T	34.548	−0.006
$\sqrt{t}$**-T**	34.549	−0.006
Shuffled	34.488	−0.004
Normalized	21.839	0.364

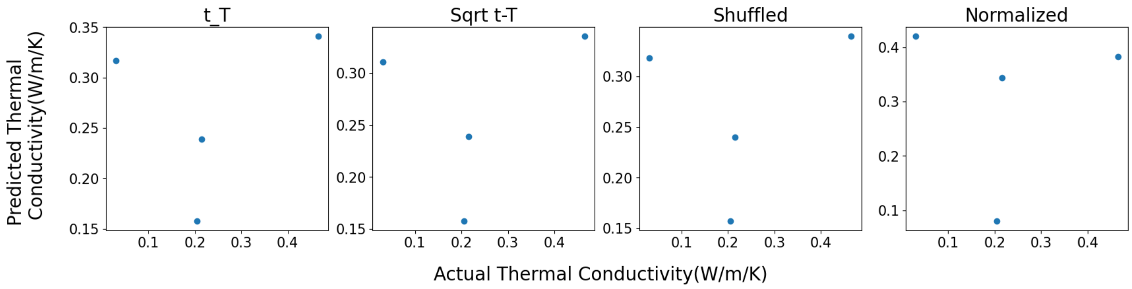

Figure 4.5: Result of training with Random Forest with data of 20 samples with a thermal conductivity lower than 1 Wm^-1K^-1.

Table 4.3: Evaluation results by dataset and regression algorithm of samples with a thermal conductivity lower than 1 Wm^-1K^-1.

	MSE	**R-squared score**
Linear Regression		
t-T	0.049	-1.042
$\sqrt{t}$-T	0.022	0.094
Shuffled	0.024	0.012
Normalized	0.155	-5.445
SVR		
t-T	0.046	-0.908
$\sqrt{t}$-T	0.038	-0.582
Shuffled	0.046	-0.908
Normalized	0.037	-0.526
Decision Tree		
t-T	0.054	-1.27
$\sqrt{t}$-T	0.055	-1.295
Shuffled	0.054	-1.27
Normalized	0.062	-1.575
Random Forest		
t-T	0.025	-0.045
$\sqrt{t}$-T	0.025	-0.023
Shuffled	0.025	-0.058
Normalized	0.048	-0.988

As a result of modeling in the group in which only samples with a thermal conductivity lower than 1 m^-1K^-1 were filtered, the MSE value is relatively small. Still, the R-squared score is negligibly tiny or negative. The result indicates that there is a significant difference between the predicted thermal conductivity value and the actual thermal conductivity value. Compared to the previous experiments, the errors shown in Figure 4.5 and Table 4.3 can be interpreted as being due to insufficient training data.

To ensure that the STD is less than 5%, the R-squared score should be high enough, and the MSE should be low enough. As a general rule of thumb, an R-squared score of 0.7 or higher and an MSE of 0.5 or lower are often considered to be acceptable for applications where the STD should be less than 5%[23], [24].

Summarizing the results of this section, the evaluation results in the group where samples with thermal conductivity lower than 50 $m^{-}1K^{-}1$ were filtered show satisfactory values (MSE:0.175, R-squared score: 0.996), it can be seen that prediction of thermal conductivity values using machine learning is valid.

4.1.2 Prediction using experimental and simulated data

Since most of the low-density/high-insulation materials that our research focuses on have thermal conductivity values in the range of 0 to $1m^{-}1K^{-}1$, Simulation data generated using COMSOL was applied in this section to compensate for the lack of data for learning and testing.

To predict the thermal conductivity of a material with a thermal conductivity lower than 1 $m^{-}1K^{-}1$, an experimental data set of 5 samples and 30 data sets generated through simulation are used. The dataset is the same as in Table 3.1, but normalized data is not used because simulation data does not use output power and sensor radius.

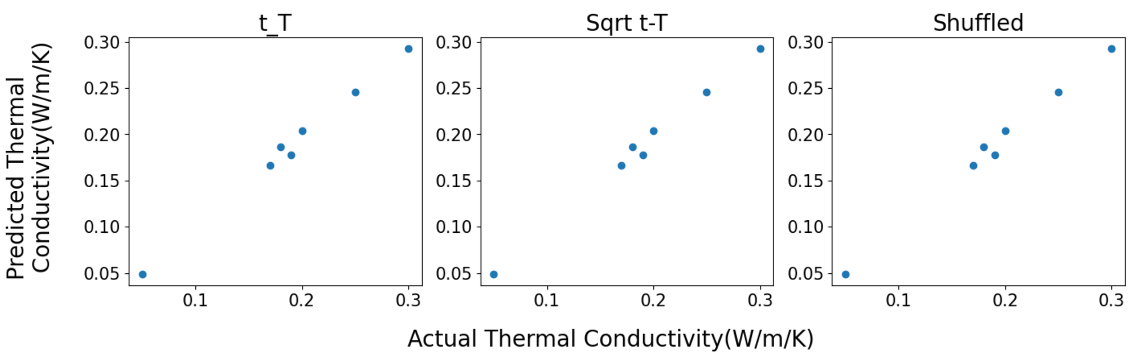

Figure 4.6: Result of training with Linear Regression with data of 35 samples with a thermal conductivity lower than 1 W$m^{-}1K^{-}1$ including simulated data.

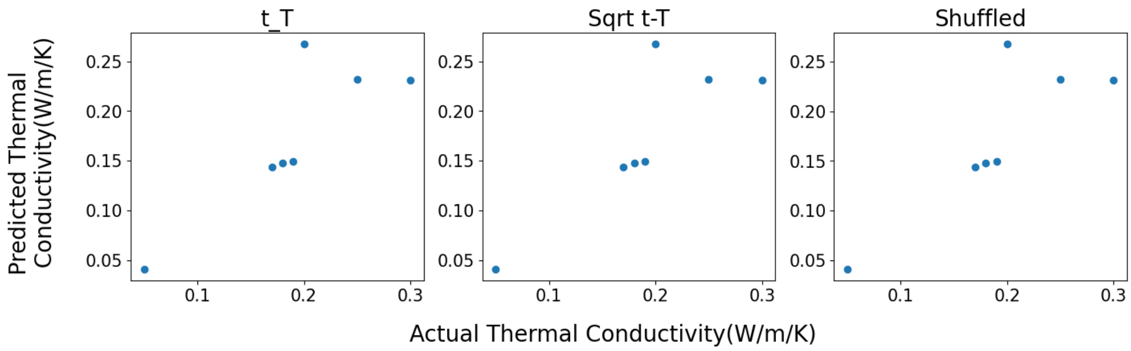

Figure 4.7: Result of training with Random Forest with data of 35 samples with a thermal conductivity lower than 1 W$m^{-}1K^{-}1$ including simulated data.

Table 4.4: Evaluation results by dataset and regression algorithm of samples with a thermal conductivity lower than 1 W$m^{-}1K^{-}1$ including simulated data.

	MSE	**R-squared score**
Linear Regression		
t-T	0.0004	0.992
$\sqrt{t}$**-T**	0.0004	0.992
Shuffled	0.0004	0.992
SVR		
t-T	0.008	−0.516
$\sqrt{t}$**-T**	0.007	−0.44
Shuffled	0.008	−0.516
Decision Tree		
t-T	0.003	0.504
$\sqrt{t}$**-T**	0.003	0.504
Shuffled	0.002	0.551
Random Forest		
t-T	0.002	0.633
$\sqrt{t}$**-T**	0.002	0.633
Shuffled	0.002	0.632

Among the evaluation results of the data set, including simulation and experimental data, the best algorithm is Linear Regression with a Shuffled dataset (MSE:3.993e-05, R-squared score: 0.992). Compensating for insufficient data, it shows remarkably accurate results compared to Table 4.3.

4.2 Part II: High-temperature measurements

4.2.1 Prediction using experimental data in various measurement temperatures

In the same approach as in Part I, the experimental data obtained through the TPS method were used to train various models for the prediction task. However, while the data of Part I was measured at room temperature, in the case of Part II, data between -10 to 1000 °C were used. In the latter part, the thermal conductivity values of the literature shown in the curve of Figure 3.3 and the predicted values were compared and analyzed.

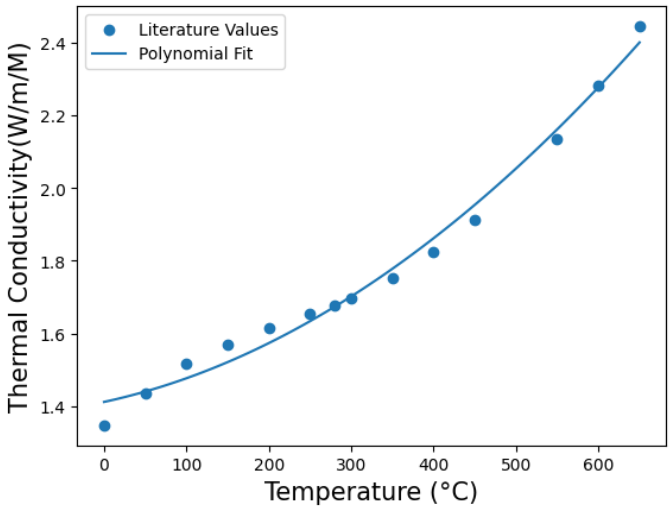

Figure 4.8: Thermal conductivity of quartz according to the measured temperature in the literature and its polynomial fit curve

Figure 4.9 shows that thermal conductivity prediction with linear regression model works very well in all data sets. Additionally, Figure 4.8 shows the results of applying polynomial fitting for comparative analysis with the values in the literature. Looking at Table 4.5, the polynomial fitting performs well (MSE:0.001, R-squared score: 0.987), but compared to the fact that the data using the TPS method is available up to 973 °C, the thermal conductivity from the literature is not provided from 650 °C or higher.

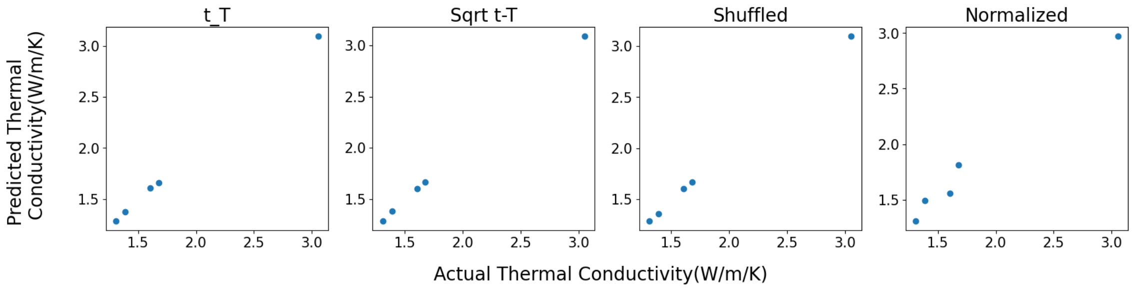

Figure 4.9: Results of training to predict thermal conductivity using linear regression and transient curves at each measurement temperature

Table 4.5: Evaluation results by dataset and regression algorithm of samples with thermal conductivity using transient curves at each measurement temperature.

	MSE	R-squared score
Linear Regression		
t-T	0.001	0.998
$\sqrt{t}$**-T**	0.0	0.999
Shuffled	0.001	0.998
Normalized	0.007	0.982
SVR		
t-T	0.016	0.96
$\sqrt{t}$**-T**	0.016	0.961
Shuffled	0.016	0.96
Normalized	0.109	0.732
Decision Tree		
t-T	0.015	0.964
$\sqrt{t}$**-T**	0.015	0.964
Shuffled	0.015	0.964
Normalized	0.015	0.964
Random Forest		
t-T	0.035	0.914
$\sqrt{t}$**-T**	0.035	0.914
Shuffled	0.035	0.913
Normalized	0.036	0.911
Polynomial Fit		
t-T	0.001	0.987

Compared with the polynomial fit of literature values in Table 4.5, the linear regression model demonstrates superior performance in terms of both mean squared errors (MSEs) and R-squared scores. This suggests that machine learning can effectively mitigate the influence of noisy transient curves. However, it is essential to note that the choice of the most suitable algorithm and its accuracy heavily relies on the temperature-dependent trend of the material's conductivity values.

5

Conclusion and Future work

5.1 Conclusion

Based on the research conducted, it can be concluded that machine learning techniques show promising results in accurately predicting thermal conductivity from transient curves in Part I of the study. However, in Part II, where noisy transient curve data generated from high-temperature measurements and literature thermal conductivity values were used, the accuracy of the predictions could be higher.

In Part I, the results demonstrated that machine learning models could accurately predict thermal conductivity from transient curves of low-density/high-insulation materials without CL value. This suggests that machine learning has the potential to be used in a complementary way without the necessary parameters, which is one of the limitations of the existing method.

In Part II, the research aimed to compensate for the difficulty of obtaining thermal conductivity values from noisy transient curve data at high temperatures. The goal was to explore whether getting thermal conductivity values with noisy transient curves through machine learning is possible. The results show the feasibility of using machine learning techniques to enhance the precision of estimating thermal conductivity values at extreme measurement temperatures. However, due to the challenges associated with noisy transient curve data and the need for reference temperature versus thermal conductivity, the accuracy of the predictions varies in the different scenarios.

Despite the potential limitations in Part II, the overall research project highlights the potential of machine learning in predicting thermal conductivity and its implications for cost reduction and improved efficiency in thermal property measurement. By leveraging machine learning algorithms, resources such as electrical energy for oven heating and machine operation can be conserved, as incorrect measurements and repeated trials are minimized.

In conclusion, while Part I of the research project demonstrated accurate thermal conductivity predictions from transient curves, Part II highlighted the challenges of predicting thermal conductivity from noisy transient curve data. Further research and refinement of machine learning models may be required to address the complexities associated with high-temperature measurements and noisy data. Nevertheless, the use of machine learning in thermal property measurement holds promise for reducing costs and improving the efficiency of the measurement process, ultimately contributing to advancements in various fields that rely on accurate thermal conductivity data.

5.2 Future work

Future research can investigate using more advanced machine learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs). These models capture intricate patterns and relationships in the data, potentially improving the accuracy of predicting thermal conductivity from transient curves.

Future studies can incorporate a broader range of sample types and expand the temperature range to enhance the effectiveness and generalization of the machine learning models. Including samples with diverse material properties and measurements at extreme temperatures can provide a more comprehensive evaluation of the models' performance and enable better prediction capabilities.

While the TPS method has shown promise in medical applications like skin cancer detection and horseshoe disease identification, further exploration and implementation of the TPS method in the medical field are warranted [25]. Researchers can focus on adapting and optimizing the TPS method specifically for medical contexts, potentially developing non-invasive diagnostic tools with improved accuracy and efficiency.

By exploring more complex machine learning models, expanding the sample types and temperature range, and investigating the potential of the TPS method in the various application fields, future work can contribute to advancements in thermal conductivity prediction from transient curves and facilitate the development of innovative applications in areas such as medical diagnostics.

Bibliography

[1] S. E. Gustafsson, "Transient plane source techniques for thermal conductivity and thermal diffusivity measurements of solid materials," *Review of scientific instruments*, vol. 62, no. 3, pp. 797–804, 1991.

[2] *Hot disk instruments publications*, `https://www.hotdiskinstruments.com/publications/`, Accessed: June 24, 2023.

[3] H. D. AB, *Hot Disk Manual.* Gothenburg, 2022.

[4] B. Mihiretie, D. Cederkrantz, M. Sundin, *et al.*, "Thermal depth profiling of materials for defect detection using hot disk technique," *AIP Advances*, vol. 6, no. 8, pp. 085–217, 2016.

[5] I. ISO, "22007-2 plasticsdetermination of thermal conductivity and thermal diffusivitypart 2: Transient plane heat source (hot disc) method," *ISO: Geneva, Switzerland*, 2015.

[6] K.-Q. Li, Q. Kang, J.-Y. Nie, and X.-W. Huang, "Artificial neural network for predicting the thermal conductivity of soils based on a systematic database," *Geothermics*, vol. 103, pp. 102–416, 2022.

[7] K.-Q. Li, Y. Liu, and Q. Kang, "Estimating the thermal conductivity of soils using six machine learning algorithms," *International Communications in Heat and Mass Transfer*, vol. 136, pp. 106–139, 2022.

[8] B. Mihiretie, D. Cederkrantz, A. Rosén, *et al.*, "Finite element modeling of the hot disc method," *International Journal of Heat and Mass Transfer*, vol. 115, pp. 216–223, 2017.

[9] Y. He, "Rapid thermal conductivity measurement with a hot disk sensor: Part 1. theoretical considerations," *Thermochimica acta*, vol. 436, no. 1-2, pp. 122–129, 2005.

[10] A. Sizov, D. Cederkrantz, L. Salmi, *et al.*, "Thermal conductivity versus depth profiling of inhomogeneous materials using the hot disc technique," *Review of Scientific Instruments*, vol. 87, no. 7, pp. 074–901, 2016.

[11] M. T, "Machine learning, new york, ny, usa: Mcgraw-hill, inc," 1997.

[12] J. H. Friedman, "Recent advances in predictive (machine) learning," *Journal of classification*, vol. 23, no. 2, pp. 175–197, 2006.

[13] X. Su, X. Yan, and C.-L. Tsai, "Linear regression," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 3, pp. 275–294, 2012.

[14] F. Zhang and L. J. O'Donnell, "Support vector regression," in *Machine learning*, Elsevier, 2020, pp. 123–140.

[15] Y.-Y. Song and L. Ying, "Decision tree methods: Applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015.

[16] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017.

[17] COMSOL, *INTRODUCTION TO COMSOL Multiphysics*. Gothenburg, 1988-2022.

[18] V. Gerlich, K. Sulovská, and M. Záleák, "Comsol multiphysics validation as simulation software for heat transfer calculation in buildings: Building simulation software validation," *Measurement*, vol. 46, no. 6, pp. 2003–2012, 2013.

[19] R. W. Pryor, *Multiphysics modeling using COMSOL: a first principles approach*. Jones & Bartlett Publishers, 2009.

[20] M. Streza, Y. Fedala, J. Roger, G. Tessier, and C. Boue, "Heat transfer modeling for surface crack depth evaluation," *Measurement Science and Technology*, vol. 24, no. 4, pp. 045–602, 2013.

[21] A. Sugawara, "The precise determination of thermal conductivity of pure fused quartz," *Journal of Applied Physics*, vol. 39, no. 13, pp. 5994–5997, 1968.

[22] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, p. 623, 2021.

[23] N. R. Draper and H. Smith, "Applied regression analysis, john wiley and sons," *New York*, vol. 407, 1981.

[24] T. Wood, "The large-scale atmospheric circulation response to climate change drivers: A multi-model comparison study," Ph.D. dissertation, University of Leeds, 2022.

[25] J. Sköld, "Detection of Damage in the Equine Hoof-A possible new application for the Hot Disk Method?", Master's thesis, Chalmers University of Technology, 2017.