

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Investigation of Gas Diffusion Layer Intrusion in PEM Fuel Cell Using Physics-informed Machine Learning

Permalink

<https://escholarship.org/uc/item/0fq8d3d3>

Author

Ly, Vu

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Investigation of Gas Diffusion Layer Intrusion in PEM Fuel Cell Using Physics-informed
Machine Learning

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE
in Mechanical and Aerospace Engineering

by

Vu Ly

Thesis Committee:
Professor Yun Wang, Chair
Professor Mark Walter
Professor Penghui Cao

2023

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iii
LIST OF TABLES	iv
ACKNOWLEDGMENTS	v
ABSTRACT OF THE THESIS	vi
1 An Introduction to Fuel Cell	1
1.1 Introduction	1
1.2 What is a Fuel Cell?	2
1.3 Fuel Cell Efficiency	4
1.4 Proton Exchange Membrane Fuel Cell	6
1.5 GDL Intrusion	9
1.6 Objective	10
2 Methods and Approach	12
2.1 Governing Equations	12
2.2 Numerical Implementation	15
2.3 Design of Experiments	17
2.4 Machine Learning	18
2.5 Optimization	25
3 Results and Discussion	27
3.1 Simulations and Two-Factorial Results	27
3.2 Results of Different Machine Learning Algorithms	33
3.3 Optimization Results	38
4 Conclusions	41

LIST OF FIGURES

	Page
1.1 Simple explanation of Fuel Cell	3
1.2 PEMFC Layout [1]	7
1.3 a) SEM image of carbon cloth (E-Tek), b) SEM image of carbon paper (Toray 060) [2]	8
1.4 Different types of gas channel configurations for BPs: (a) straight parallel; (b) interdigitated; (c) pin-type; (d) spiral; (e) single-channel serpentine; (f) multiple-channel (triple) serpentine [3]	9
1.5 Fuel Cell Assembly [4]	10
1.6 GDL intrusion in GFC at four different pressures [5]	11
2.1 Schematic of BP and GDL	16
2.2 Plot showing the performance vs. Interpretability of each machine learning algorithm [6]	20
2.3 Decision Tree	21
2.4 Plot explaining Support Vector Machine [7]	22
2.5 Plot explaining K-Nearest Neighbor with K = 3 [7]	23
2.6 Plots demonstrating underfitting, good fit, and overfitting [6]	24
2.7 Schematic diagram of the steps for this experiment	26
3.1 BP and GDL model from COMSOL software	28
3.2 Mesh size of 2×10^{-5} m for the model	28
3.3 Schematic of BP and GDL	29
3.4 Comparison of simulation results with literature data	30
3.5 Half the intrusion area of the GDL into the channel	31
3.6 F and P values of the five parameters and relationship represented in a half-normal plot	32
3.7 Effects of different parameters on the intrusion area	34
3.8 Results of each model, a) Linear Regression, b) Decision Tree, c) KNN, d) SVR	37
3.9 Optimization results for a) Pressure vs GDL Height and b) GDL Young's Modulus vs Rib Configuration	39
3.10 Optimization results for a) Channel Width vs GDL Height and b) GDL Height vs Pressure	40

LIST OF TABLES

	Page
1.1 Characteristics of each type of fuel cell	3
1.2 Typical PEMFC operating conditions	7
2.1 Material properties of BP and GDL	17
2.2 Inputs for two-factorial design	18
3.1 Inputs for two-factorial design	27
3.2 F-values and P-values obtained from Design-Expert software	32
3.3 Inputs for the parametric study	33
3.4 Models accuracy and RMSE	37

ACKNOWLEDGMENTS

I want to thank my advisor, Prof. Yun Wang, for all his time, guidance, and support over this past year. I am incredibly grateful that he reached out to me and allowed me to work on subjects that I found interesting. I appreciate all our research discussions, and he always encourages me to pursue topics that will help me in the future.

I want to thank my Master's committee members, Prof. Mark Walter and Prof. Penghui Cao, for taking time out of their busy schedules to serve on the committee.

I was incredibly fortunate at UC Irvine to have my friends Nareg Haladjian, Omid Souri, Andre Necochea, Pablo Muro, Co Ly, Nghi Nguyen, and Trung Tran, who were always there to turn bad days into good, hunger into delicious meals, and almost-tears into belly laughs. I have also had incredible friends from community college, including Jeffrey Tsang, Sim Kay, and Holly Rheinschild. I have shared unforgettable travel experiences, laughs, joy, and support. These fantastic individuals have helped shape me into who I am today, and I will always be grateful.

Finally, I would like to thank my parents, John Ly and Loan Ly, for their unconditional love and support throughout my life and their sacrifice for moving to a new country so my siblings and I have a better future. I have had incredible support from my siblings, Johnny Tran, Jason Ly, Truong Ly, and Ace Ly, especially my older sister Cherilyn Ly who has always been a model of hard work and perseverance to look up to and a soft and understanding voice to fall back on. Their wholehearted support is what motivates me to do my best every day. I am grateful to have them in my life; none of my success would have been possible without them.

ABSTRACT OF THE THESIS

Investigation of Gas Diffusion Layer Intrusion in PEM Fuel Cell Using Physics-informed Machine Learning

By

Vu Ly

Master of Science in Mechanical and Aerospace Engineering

University of California, Irvine, 2023

Professor Yun Wang, Chair

This research investigated the gas diffusion layer (GDL) membrane intrusion into a gas flow channel (GFC) using physics-informed machine learning for PEM fuel cells. This study was done to reduce the time it takes to simulate the model and find the intrusion area. To establish the training data for machine learning, different configurations of the GFC and GDL were created using COMSOL to simulate the GDL intrusion into a gas flow channel. The data from these simulations then get exported into Design-Expert, a statistical software to determine the parameters with the highest impact on the % intrusion. The GDL's Young's Modulus (E_{GDL}) was found to be the parameter with the most significant impact, with an F-value of 3776.66 compared to the next parameter Bipolar Plate Channel Width (CW_{BP}) with an F-value of 992.59, or the lowest parameter GDL Height (H_{GDL}) with an F-value of 159.96. Design-Expert also shows that E_{GDL} and the Bipolar Plate Rib (R_{BP}) positively affect the % intrusion area, where both will yield lower % intrusion the higher the parameters are. Whereas higher the values are CW_{BP} , H_{GDL} , and Pressure (P) will increase the % intrusion.

With the finding from the Two-Factorial test, the simulation interval for E_{GDL} is higher than the other parameters to yield more accurate training data for machine learning. Then performing, a parametric study will be done to find the x-y coordinates of the intrusion curve to export to

MATLAB to find the intrusion area. Four machine-learning algorithms, linear regression, Decision Tree, SVR, and KNN, were deployed to train using 70% of the data set and the remaining 30% for testing. The accuracy of each model were calculated based on how close the prediction is compared to the actual value. It was found that out of the four algorithms, Linear Regression has the lowest model accuracy at 68.5% and the highest RMSE at 0.0751, and Decision Tree has the highest model accuracy at 95.5% and lowest RMSE at 0.0303. Thus, Decision Tree was used to make predictions for various ranges of the five parameters to find the optimization parameters for design. This study was done to help reduce the time it takes to simulate the model and find the intrusion area. It takes an average of 20 minutes to create the model, simulate it, and calculate to find the intrusion. With the trained machine learning models, the intrusion area can be found in less than a minute. The machine learning model also identifies the parameter ranges for less than 10% and 20% intrusion to guide fuel cell material selection and design.

Chapter 1

An Introduction to Fuel Cell

1.1 Introduction

Fossil fuels, such as coal and oil, release a large amount of carbon dioxide into the atmosphere after they get burned. This significantly contributes to climate change and increases global temperatures, which negatively affect the environment and the economy, including the degradation of ecosystems, the loss of agricultural land, and rising sea levels. Studies have shown that these effects could cost the global economy up to 0.5 trillion dollars annually [8], [9].

To combat these threats, the European Union has committed to reducing its greenhouse gas emission by 80-95% by 2050 [10], which has further motivated researchers to find a better solution for renewable and sustainable energy systems. While many different renewable energy sources are available, such as solar and wind power, their intermittent nature has led to a search for more efficient and reliable options. Began in the 1970s, fuel cells were introduced and have gained popularity as an alternative power source and are seen as a potential replacement for fossil fuels. Due to their high energy efficiency, and low pollutant emission, the fuel cell became a prime candidate for this mission. It directly converts the chemical energy of a fuel, such as

a hydrogen, into electricity through a chemical reaction where water and heat are the byproducts, making it a clean and efficient power source. This process is called electrolysis, and the hydrogen resulting from the process is considered green because it does not produce greenhouse gas. Additionally, unlike renewable energy sources such as solar and wind power, fuel cells can provide a consistent and reliable source of energy, making them a promising option for meeting the increasing demand for clean and sustainable energy.

1.2 What is a Fuel Cell?

A fuel cell converts chemical energy like hydrogen into electricity through a chemical reaction—Figure 1.1 shows a simple box with an input and an output, where the inputs are hydrogen and oxygen being fed into the fuel cell at the anode and cathode. From the output side, these chemical reactions' byproducts are water, heat, and electricity, making them clean and efficient energy sources. The electricity produced by the fuel cell can be used to power various devices, including cars, portable electronics, and stationary power systems. The heat produced by the reaction can also be harnessed to provide additional power or to heat buildings and water [11]. Due to the wide range of applications and benefits, the development of fuel cells is an active area of research. Researchers hope to improve fuel cells' efficiency and cost-effectiveness to make them a more viable and sustainable energy source in the future.

Researchers and engineers developed several types of fuel cells [12]. These types include solid oxide fuel cells (SOFC), molten carbonate fuel cells (MCFC), and phosphoric acid fuel cells (PAFC). Among these types, polymer electrolyte membranes (PEM) are the most popular due to their long lifetime, lower manufacturing costs and lower operating temperatures [12]. The characteristics of each type of fuel cell are shown in Table 1.1 [12].

These fuel cells have different operating temperatures and efficiencies, and researchers are

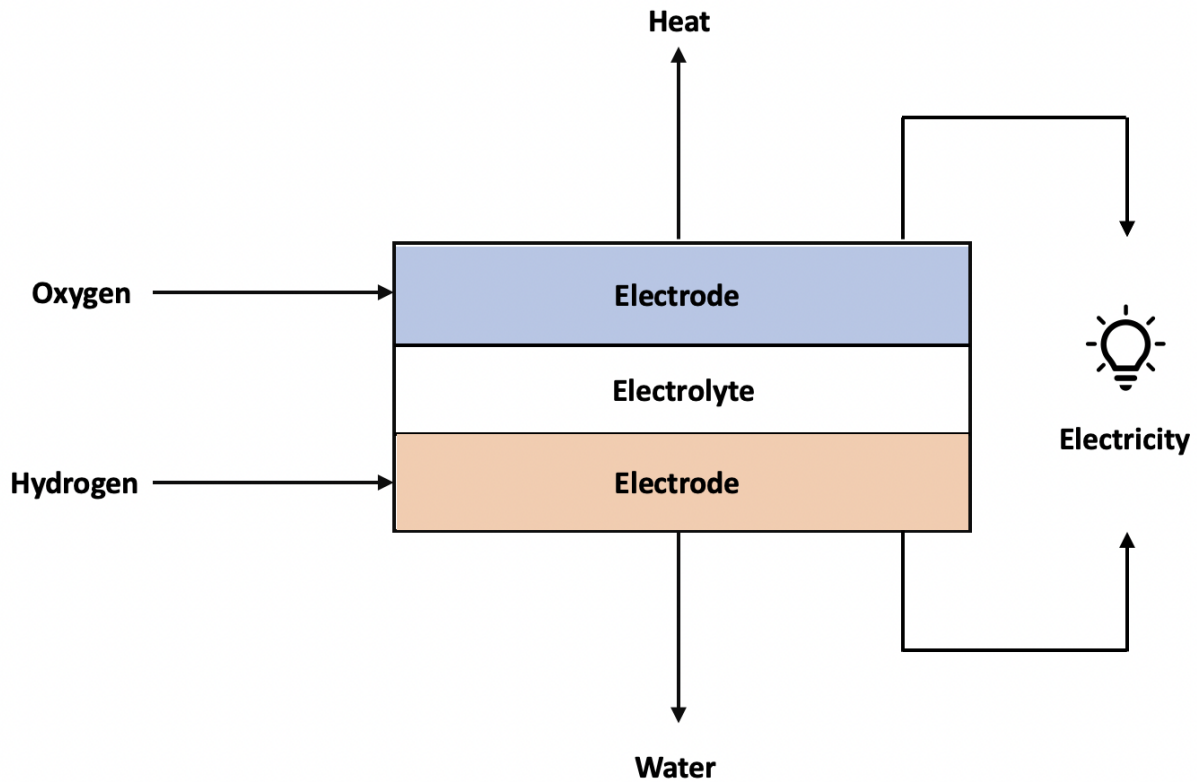


Figure 1.1: Simple explanation of Fuel Cell

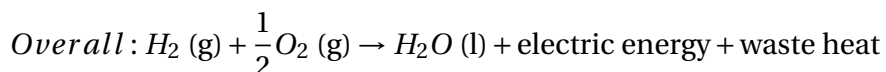
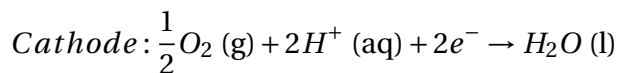
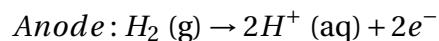
Table 1.1: Characteristics of each type of fuel cell

Fuel Cell Type	Operating Temperature (°C)	Electrical Efficiency (%)	Applications
SOFC	600 – 1000	35 - 43	Electric utility
MCFC	600 - 700	45 - 47	Electric utility
PAFC	150 - 200	> 40	Distributed generation
PEMFC	50 - 100	53 - 58	Transportation

working to develop applications that take advantage of these characteristics. In fact, due to the low operating temperatures of the PEMFC and environmentally friendly byproducts, it is being commercialized in the automotive industry. In 2018, Japan produced 2,459 fuel cell vehicles, and the country plans to increase production to 800,000 by 2030 [13].

The electrochemical reactions that occur on the anode and cathode of the fuel cell are shown

in the equations of the chemical reactions below [14].



The system's byproducts are heat and water, as seen in the overall reaction. It is essential to continue removing these byproducts from the system because their accumulation can reduce the fuel cell's lifespan and decrease its efficiency. For example, the fuel cell can overheat and lead to thermal damage if the excess heat is not removed, or the water build-up can flood the inside and corrode the components. Studies have shown that water management is essential to achieve high performance and durability in a PEM fuel cell [15], [16]. To prevent these problems, fuel cells are designed with mechanisms for removing the water and heat byproducts. For example, some fuel cells use water-management systems that circulate coolant through the cell to remove excess heat [17] and prevent flooding. The heat can also be transported through conduction in the solid phase or through convection and conduction in the fluid phase [14]. Other fuel cells use mechanical systems or passive mechanisms to remove the water and heat byproducts.

1.3 Fuel Cell Efficiency

Fuel cell efficiency is generally higher than internal combustion engines because fuel cells directly convert chemical energy into electricity. In comparison, internal combustion engines will convert chemical energy into heat and mechanical energy. This means that some of the energy from the fuel is lost as heat, which lowers the engine's overall efficiency. This is one of the reasons why internal combustion engine often has an efficiency of 30% or lower [18]. On the other

hand, PEMFCs, have a theoretical efficiency as high as 80% at room temperature. This means that about 80% of the chemical energy from the fuel is converted into usable electricity, with the remaining 20% being lost as heat or other forms of waste. Equation 1.1 below shows the theoretical efficiency of the PEMFC at room temperature:

$$\mu = \frac{\Delta G}{\Delta H} = \frac{237.34}{286.02} = 83\% \quad (1.1)$$

Where ΔG is the Gibbs free energy and ΔH is the enthalpy difference between the products and reactants. Gibbs's free energy is a measure of the potential for a chemical reaction to occur, and enthalpy is the total heat content of the system. Both values are temperature dependent, and Gibbs free energy will decrease faster than enthalpy as the temperature increase. This means the efficiency of PEMFC is lower at a higher temperature. For this reason, PEMFC operates between 50 - 100°C. However, the actual efficiency of a PEMFC in real-world conditions is lower than this theoretical value due to various factors, such as activation and ohmic losses. Additionally, the efficiency of a PEMFC can be reduced by the production of entropy, which is a measure of the disorder or randomness in a system, and it is produced when chemical reactions occur in the fuel cell. Entropy is being considered when calculating the theoretical potential of a PEMFC using the change in Gibbs free energy, ΔG , change in enthalpy, ΔH , and temperature as shown in equation 1.2 below:

$$\Delta G = \Delta H - T\Delta \quad (1.2)$$

In the case of a PEMFC, the potential can be calculated by rearranging the equation above and substituting the appropriate values. For example, suppose that the reaction occurs at a room temperature of 25 degrees Celsius and a pressure of 1 bar; the potential can be calculated as seen in equation 1.3 [19]:

$$E = \frac{-\Delta G}{nF} = \frac{237340}{(2)(96485)} = 1.23V \quad (1.3)$$

Where F is Faraday's constant (the amount of electric charge per mole of electrons), and n is the number of electrons per molecule of hydrogen, which is 2. The theoretical potential of a PEMFC is an important parameter that determines the efficiency of the fuel cell and the amount of electricity it can generate. Higher potential values indicate a more significant potential for the fuel cell to generate electricity, while lower potential values indicate a lower potential for electricity generation. With the current technologies, the peak PEMFC efficiency is about 65%, which is expected to reach 70% in the near future [20].

1.4 Proton Exchange Membrane Fuel Cell

A polymer electrolyte membrane fuel cell (PEMFC) is a type of fuel cell that uses a proton exchange membrane as the electrolyte, as shown in Figure 1.2, [1]. A typical PEMFC operates under specific conditions, such as a certain pressure, temperature, and flow rate of hydrogen, oxygen, and air, as shown in Table 1.2 [21]. The humidity of the reactants is also carefully controlled to ensure optimal fuel cell performance. As the name implies, the heart of the fuel cell is a membrane separating hydrogen and oxygen, preventing electrons from passing through and only allowing protons to pass through via diffusion [21]. A common material used for the membrane is DuPont Nafion®, and it is the most reliable ionomer used for PEMFC [22]. The left side of the PEMFC consists of negatively charged electrodes, where hydrogen is oxidized, while the right side consists of positively charged electrodes, where oxygen is oxidized. When hydrogen and oxygen come in contact with the electrolyte, they react to form water and release energy in the form of heat and electricity [23].

Sandwiching between the membrane of the PEMFC are the catalyst layers (CLs), which usually

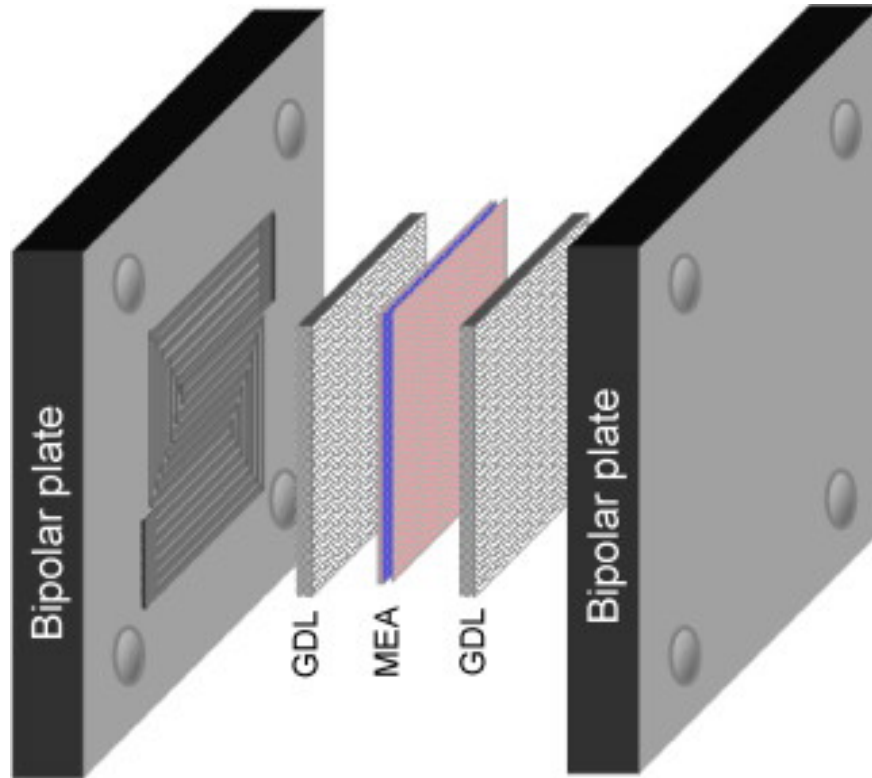


Figure 1.2: PEMFC Layout [1]

Table 1.2: Typical PEMFC operating conditions

Pressure	H_2/O_2 : 1 - 3 bar
Temperature	$20^\circ C - 90^\circ C$
Stochastic ratio	H_2 : 1.2-1.5 O_2 : 1.5 - 2
Relative Humidity of reactants	H_2 : 0% to 100% O_2/Air : 0% to 100%

use platinum or its alloys as the catalyst. The purpose of the CLs is to catalyze the electrochemical reaction that occurs in the fuel cell, allowing it to produce electricity, heat, and water at low temperatures [21]. In figure 1.2, the CL is part of the Membrane Electrode Assembly (MEA). Next to the CLs is the gas diffusion layer (GDL), which is usually based on carbon paper or carbon cloth, as shown in Figure 1.3 [2]. The GDL is a porous structure that is in the millimeter range, allowing the flow of reactant gases and water. The GDL porosity is typically 60-80% open to ensure sufficient gas flow [24]. The CLs and GDLs are essential components of a PEMFC, as

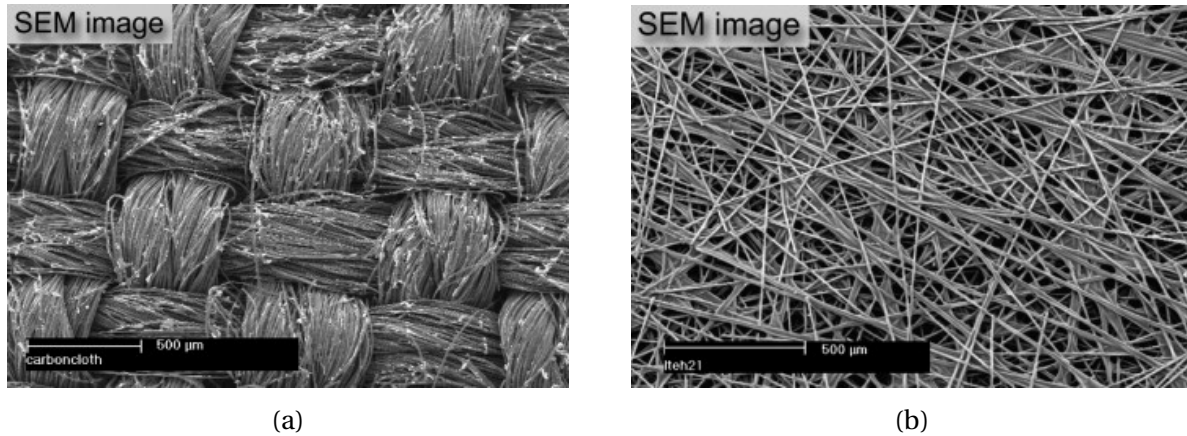


Figure 1.3: a) SEM image of carbon cloth (E-Tek), b) SEM image of carbon paper (Toray 060) [2]

they facilitate the electrochemical reaction and enable the fuel cell to produce electricity, heat, and water efficiently. The structure and composition of these layers are carefully designed to optimize fuel cell performance.

The bipolar plates (BPs) in a PEMFC are two plates that provide gas flow channels to distribute gas reactants and thermal/electrical pathways for heat removal, and current collection [25]. The BPs are used to separate different fuel cell stacks in a multi-cell configuration. The BPs are exposed to a corrosive environment with a pH range of 2-3 and temperatures between 60-80 degrees Celsius [26]. Because of this, BPs are often made of graphite, polymer blended with carbon, and metal and are a significant contributor to the weight and cost of a PEMFC [27]. They account for 80% of the total weight of the fuel cell and 45% of the stack manufacturing cost [26]. As such, it is essential to carefully design and select materials for the BPs to ensure they are durable and cost-effective. Effective BP design can increase the performance of a PEMFC by nearly 50% [28], making it an important area of study. Several studies have been conducted on the different types of gas channel configurations, as shown in Figure 1.4 [3]. Additionally, Shimpalee assessed the geometry of the gas channels, such as their length, width, or height, and has been studied to determine its impact on PEMFC performance [29].

This thesis focuses on the GDL of PEMFC, an essential fuel cell component that provides liquid

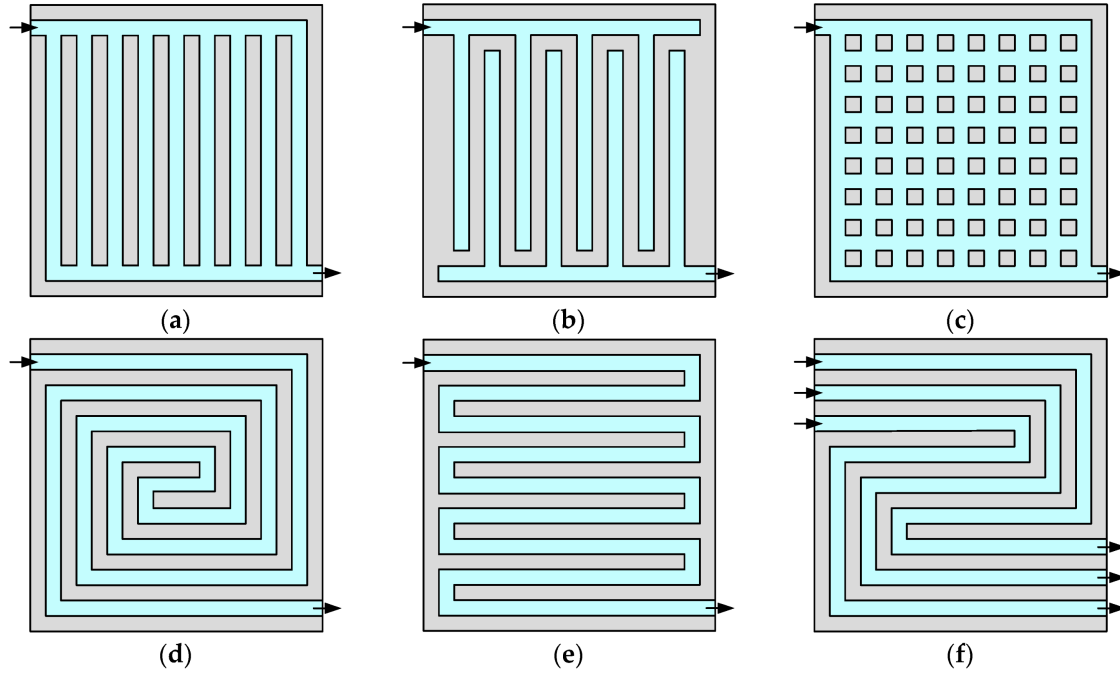


Figure 1.4: Different types of gas channel configurations for BPs: (a) straight parallel; (b) interdigitated; (c) pin-type; (d) spiral; (e) single-channel serpentine; (f) multiple-channel (triple) serpentine [3]

water, heat removal, electric conductance, and reactant transport [4]. It is a porous layer, which is subject to deformation and then intrusion into the gas flow channel upon BP compression.

1.5 GDL Intrusion

During fuel cell assembly, the GDL, membranes, BPs, and gasket are clamped together using bolts as shown in Figure 1.5 [4], [30], [31]. BPs will impose compressive stress on the GDL, which causes it to deform and intrude into the gas flow channels (GFC), whose purpose is to remove products from PEMFC and distribute fuel to the whole cell. The GDL intrusion will reduce the channel space for gas distribution, raising the gas flow pressure drop and pumping power and risks of flow maldistribution [32], [33]. A 5% intrusion of the GDL into the GFC can reduce the reactant flow inside the channel by 20% [5]. Chi-Hui Chien found that when bolt loading onto the GDL reaches 7 MPa, it will reduce the volume flow of the GFC by 8.35%

and decrease the efficiency of the PEMFC [34]. Wei Sun found that compression in GDL will reduce the porosity but increase the electronic conductivity [35]. Bazylak's research illustrates the degradation of the GDL under varying compression pressures [36]. Figure 1.6 [5] shows the intrusion at different pressures exerted on the top face of the BP. As expected, the intrusion increases as pressure increases.

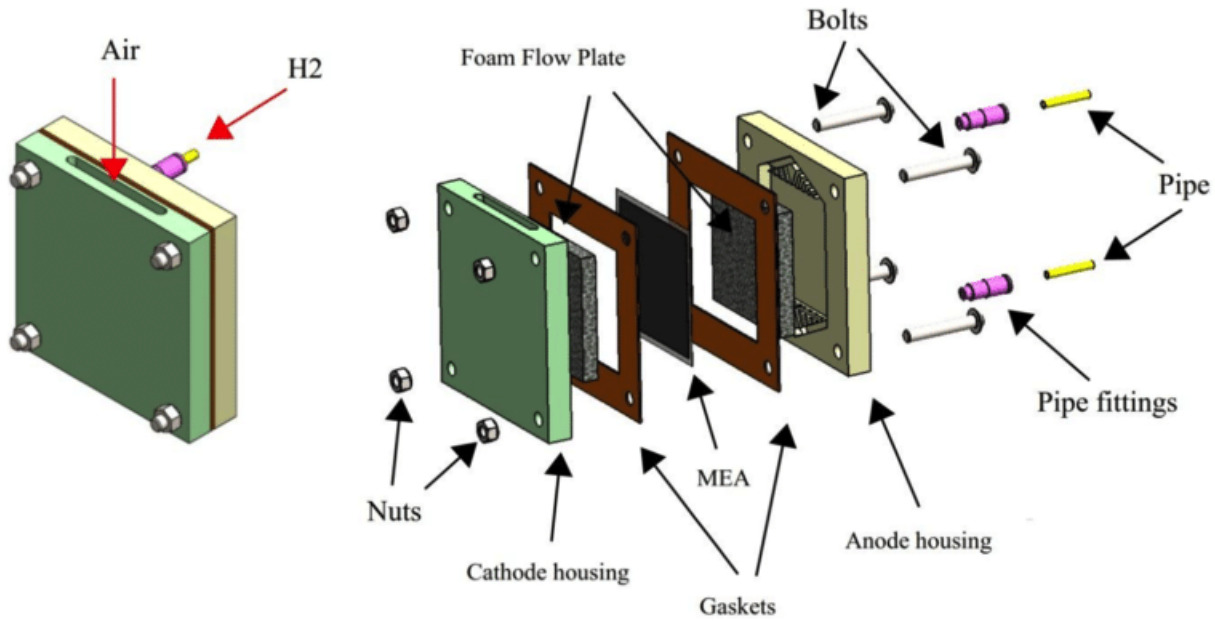


Figure 1.5: Fuel Cell Assembly [4]

1.6 Objective

The GDL intrusion into the GFC of a PEMFC can significantly reduce the reactant flow and decrease the efficiency of the fuel cell. Alex found that the contact resistance between the BP and GDL degrades the PEMFC performance [37]. Lianhong found that 59% of the total power loss in PEMFC is due to the contact resistance [38]. By performing parametric studies and using simulation and machine learning techniques, it is possible to investigate and predict the effects of different GDL and rib parameters on the intrusion of the GDL into the GFC. The parametric studies will involve varying parameters related to the GDL and rib configuration, such as the

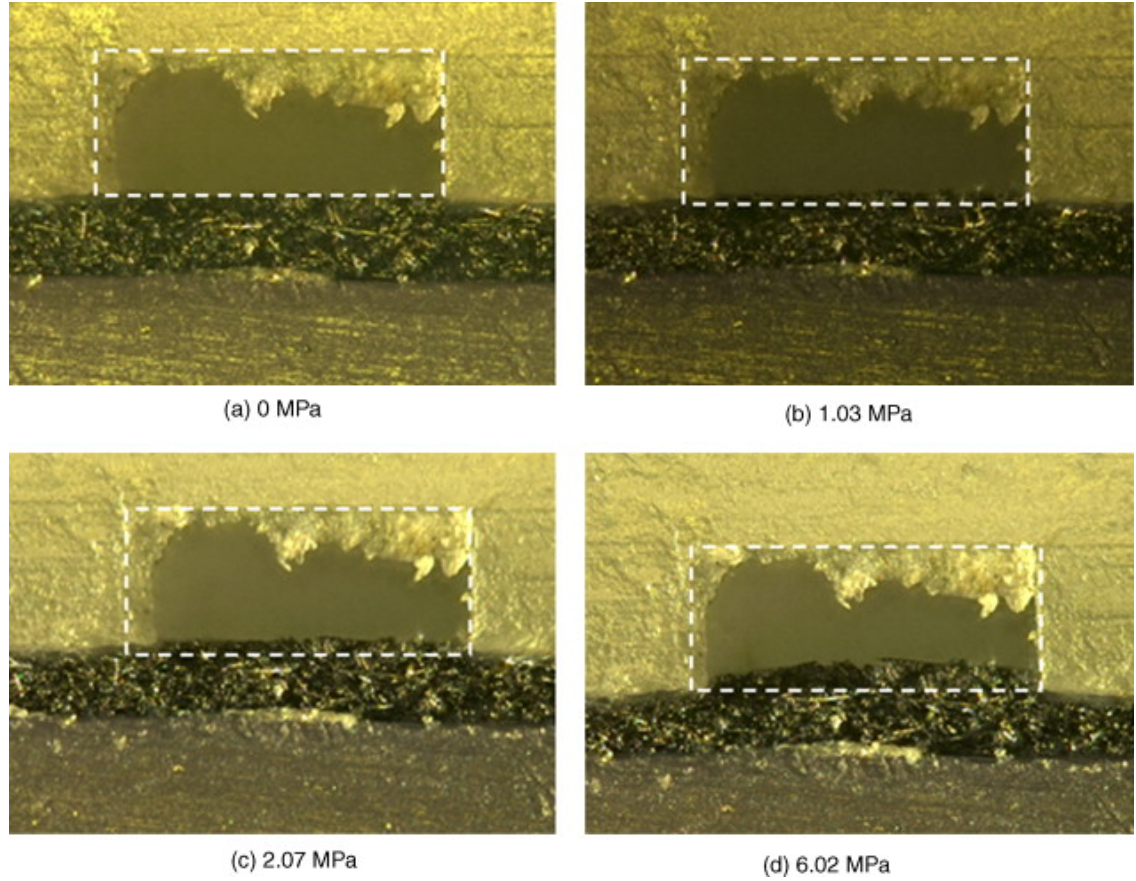


Figure 1.6: GDL intrusion in GFC at four different pressures [5]

material, thickness, and composition, as well as the geometry and loading of the BP. These studies will help to identify the parameters that have the most significant impact on GDL intrusion.

The results of the simulations can then be used to train different regression machine learning algorithms, such as Linear Regression, SVR, KNN, and Decision Trees. These algorithms can learn the relationships between the rib and GDL parameters and the intrusion of the GDL to make predictions for new input parameters. Finally, using these trained machine learning models, recommend the best rib and GDL parameters for different inputs to minimize the intrusion of the GDL into the GFC. This can help improve the performance of PEMFCs and make them more practical for real-world applications.

Chapter 2

Methods and Approach

2.1 Governing Equations

The 2D model of the rib configuration and GDL in the PEMFC is created and analyzed using the COMSOL software's static simulations. Static simulations are used to study the behavior of a system at an equilibrium, where the forces and deformations are independent of time. For this thesis, the simulations will examine how the GDL deforms under the clamping pressure produced by the bolts. It will analyze the deformation of the GDL into the channel by using the Lagrangian formulation, which tracks the body as it rotates, translates, and deforms. The formulation is built upon the continuum assumptions, which assume the materials to be homogeneous, isotropic, continuous, and independent of any particular coordinate system [39]. The deformation of the body or strain is created when it experiences external force applied to it. The body will try to resist the external force by developing the internal force or stress. The material will fail if the stress exceeds the material properties. The deformation of the body can be measured by the second-order tensor, deformation gradient, F , as shown in equation 2.1 below:

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \mathbf{I} + \frac{\partial \mathbf{u}}{\partial \mathbf{X}} = \mathbf{I} + \nabla \mathbf{u} \quad (2.1)$$

Where \mathbf{X} is the original location of the material particle, \mathbf{x} is the new location after a certain time, t , \mathbf{u} is the displacement vector, and \mathbf{I} is the identity tensor. \mathbf{X} and \mathbf{x} are in two different coordinate systems, where \mathbf{X} is in the material frame, which is fixed to the body, and \mathbf{x} is in the spatial frame, which is fixed in space. The displacement, \mathbf{u} , provides the transformation from the material frame to the spatial frame through equation 2.2 shown below:

$$\mathbf{x} = \mathbf{X} + \mathbf{u} \quad (2.2)$$

The deformation gradient, \mathbf{F} , contains information about the material's rotation and deformation. Which can be separated into either pure rotation or deformation using the Polar decomposition theorem as shown in equation 2.3 below [39]:

$$\mathbf{F} = \mathbf{R}\mathbf{U} \quad (2.3)$$

Where \mathbf{U} is the right stretch tensor, and \mathbf{R} is the pure rotation matrix \mathbf{R} . In this equation, the deformation is described first, then followed by the rotation. On the other hand, the deformation gradient can also be decomposed into equation 2.4:

$$\mathbf{F} = \mathbf{V}\mathbf{R} \quad (2.4)$$

Where \mathbf{V} is the deformation described by the left stretch tensor after the body rotation. From these two stretch tensors in equation 2.3 and 2.4, it is possible to compute the deformation of the material without knowing the rotation matrix, as shown in equation 2.5 and 2.6 below:

$$\mathbf{F}^T \mathbf{F} = (\mathbf{R}\mathbf{U})^T \mathbf{R}\mathbf{U} = \mathbf{U}^T \mathbf{R}^T \mathbf{R}\mathbf{U} = \mathbf{U}^T \mathbf{U} = \mathbf{U}^2 = \mathbf{C} \quad (2.5)$$

$$\mathbf{F}\mathbf{F}^T = \mathbf{V}\mathbf{R}(\mathbf{V}\mathbf{R})^T = \mathbf{V}\mathbf{R}\mathbf{R}^T \mathbf{V}^T = \mathbf{V}\mathbf{V}^T = \mathbf{V}^2 = \mathbf{B} \quad (2.6)$$

Where \mathbf{C} is the right Cauchy-Green deformation tensor, and \mathbf{B} is the left Cauchy-Green deformation tensor. Both also describe the deformation in two different coordinate systems: \mathbf{C} is in the material coordinate, and \mathbf{B} is in the spatial coordinate system. From these two tensors, it is possible to compute the principal stretches or eigenvalues to determine the material's change in length or volume. However, the software COMSOL will solve the deformation by using \mathbf{C} rather than \mathbf{B} , and the change in volume can be found as shown below in equation 2.7:

$$\frac{V}{V_0} = \lambda_1^2 \lambda_2^2 \lambda_3^2 \quad (2.7)$$

Prior to any rotation, the Green-Lagrange strain tensor, \mathbf{E} , can be calculated by using equation 2.8:

$$\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{I}) \quad (2.8)$$

Or using the displacement components and in Cartesian coordinates in equation 2.9 below:

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} + \frac{\partial u_k}{\partial X_i} \frac{\partial u_k}{\partial X_j} \right) \quad (2.9)$$

The Green-Lagrange strain tensor is important in finite element analysis because it computes the strain at different points inside a material. Then the result at each individual point is then assembled to obtain the overall strain distribution of the body. The strain is a measurement of the change in the shape of the material that occurs after experiencing an external load. In finite

element analysis, it is used in conjunction with the Cauchy stress tensor to analyze the behavior of the materials. The Cauchy stress tensor is a symmetric tensor that describes the stress state of any point in the body, and it is defined in equation 2.10 [39]:

$$\mathbf{t}_n = \mathbf{T}\mathbf{n} \quad (2.10)$$

Where t_n is the stress vector, T is the stress tensor, and n is the unit normal vector. Similar to the Green-Lagrange strain tensor, the Cauchy stress tensor is computed at each element in the body and then assembled at the end to determine the stress distribution within the body. Together, these two tensors will calculate the deformation and stress the GDL is experiencing in COMSOL.

2.2 Numerical Implementation

Using COMSOL, a single channel of fuel cell with two lands, which are the ridges present on both sides of the BP and GDL was developed, similar to Figure 2.1. The bottom surface of the GDL was fixed to constrain all degrees of freedom to prevent it from displacing. A positive assembly pressure P is applied on the top boundary of the BP. Lee found that performance of the PEMFC is a function of the compression pressure and GDL materials [40]. Ge, Escibano, Zhou, and Wu found that fuel cell performance decreases as the compression force increases [41], [42], [43]. Khajeh studied different parameters and how varying them will affect the efficiency of the PEMFC [44]. Nitta found that the GDL compressed very little under the channel, but the GDL under the lands is compressed to the gasket thickness [45]. This study aims to vary five different parameters in the model: the pressure (P), the bipolar plate channel width (CW_{BP}), the bipolar plate rib (R_{BP}), the gas diffusion layer height (H_{GDL}), and the Young's Modulus of the GDL (E_{GDL}) and see which parameter will affect the PEMFC the most.

Both the BP and GDL were modeled as a linear elastic, isotropic material with a mesh size of 0.02mm. To make sure the mesh size will not affect the result greatly, a study varied the sizes ranging from 50% to 200% of the employed mesh size to find the intrusion height only varied by less than 1.5% [5]. This suggests that the mesh size of 0.02mm is sufficiently small to represent the materials' behavior accurately. The BP is made of carbon graphite, and the GDL is made of carbon paper. Their material properties are shown in Table 2.1.

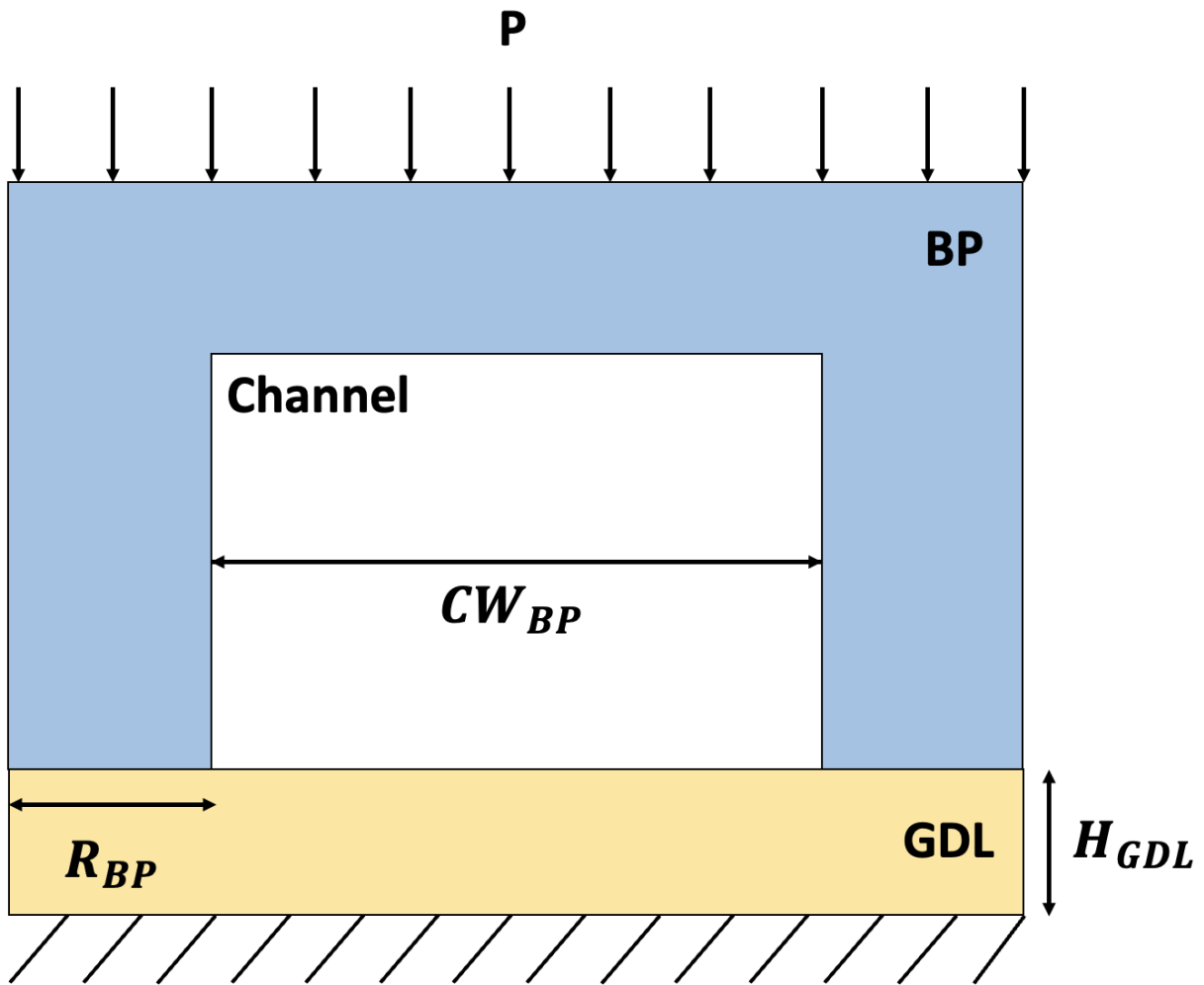


Figure 2.1: Schematic of BP and GDL

Table 2.1: Material properties of BP and GDL

Parameters	Value
BP Young's Modulus [GPa] [46]	185
BP Poisson's Ratio [37]	0.29
GDL Young's Modulus [GPa] [47]	6.3
GDL Poisson's Ratio [47]	0.09

2.3 Design of Experiments

Before performing parametric studies to gather training data for machine learning, it is necessary to evaluate whether each chosen parameter will affect the PEMFC and determine the primary parameter that affects the PEMFC. This can be done using the two-factorial design process in Design-Expert, a software package that is used exclusively on experimental design [48]. The software will use different statistical techniques to determine the effects of multiple variables on a particular output by calculating the P and F values. The P-value is the measure of the probability that the result occurred by chance. A P-value of less than 0.05 (5%) is commonly used to determine whether or not the result is statistically significant. This means that there is less than a 5% chance that the result happened by chance or 95% confidence that the result is significant [49]. On the other hand, the F-value is used to determine whether the result is significant [49]. A higher F-value indicates that the parameter has a stronger relationship with the result and, therefore, has a greater impact on the result. These two values can then be used to determine the interval of each parameter for the parametric studies. The parameters that affect the output the most will have a more extensive interval than those that do not. This is to help ensure that the models are accurate and effective at predicting the results.

As the name two factorial implies, each parameter will have two values, a low and a high, as shown in Table 2.2 [50]. The purpose is to determine each parameter's effect over a range of values. In addition, each parameter will be classified as either continuous or categorical. A continuous variable can take on any value within a given range; for example, the pressure range

is continuous because it can take on any value. A categorical is one with each value representing a different category; for example, the BP can be either graphite, polymer blended with carbon or metal. All five parameters are continuous in this study. The experiment will require 2^k runs, where "k" is the number of parameters. Thus, our study will require 32 different results that vary for each of the five parameters. After completing the two-factor experimental design study, the Design-Expert software will generate various plots and data to identify the parameter that has the most significant impact on the performance of the PEMFC and the parameter that does not affect the result, if there is one. These results can be used to design the range of values for each parameter in the study's machine-learning portion.

Table 2.2: Inputs for two-factorial design

Input	Type	Low	High
H_{GDL} [mm]	Cont.	0.2	0.6
E_{GDL} [MPa]	Cont.	6.3	31.5
CW_{BP} [mm]	Cont.	1.0	1.5
R_{BP} [mm]	Cont.	0.5	1.0
P [MPa]	Cont.	0.5	5.0

2.4 Machine Learning

Similar to the way machine learning has been widely applied in many different fields, varying from something as simple tasks like creating an email spam filter to complex ones like predicting stock market trends or to image detection of cancer. Many researchers in PEMFC have also started integrating machine learning to analyze and improve the performance of fuel cell technology [51], [52], [53]. Kannan used the Monte Carlo machine learning approach to achieve 90% accuracy in predicting the performance of fuel cells [54]. Saco used different algorithms to enhance the performance of PEMFC by varying different humidification processes [55]. Machine learning algorithms use many different mathematical models to learn from a large volume of

training data, make predictions, find patterns, and classify data or optimize based on given parameters. There are two different types of machine learning, and each has different applications and usage:

- **Supervised Learning** - the algorithm has access to a labeled set of inputs and outputs; then, it will be able to make predictions based on these data.
 - **Regression** - make predictions on the continuous variables, as mentioned in section 2.3
 - **Classification** - classify data sets to their correct category
- **Unsupervised Learning** - the algorithm does not have access to labeled data or output. Instead, it learns the patterns from the data to understand the relationship between each parameter
 - **Clustering** - grouping unlabeled data based on the relationships that appear in the data set
 - **Association** - finding the relationship between different parameters within the data set
 - **Dimensionality reduction** - reduce the dimension of the data while maintaining the integrity of the data set

In this study, supervised regression learning will be used to make predictions of the intrusion area into the GFC. Four different algorithms, Linear Regression, Decision Tree, Support Vector Regressor (SVR), and K-nearest neighbors (KNN), will be explored to find the algorithm with the highest accuracy and lowest Mean Square Error (MSE). The performance and interpretability of these different machine learning algorithms can be seen in Figure 2.2 [6].

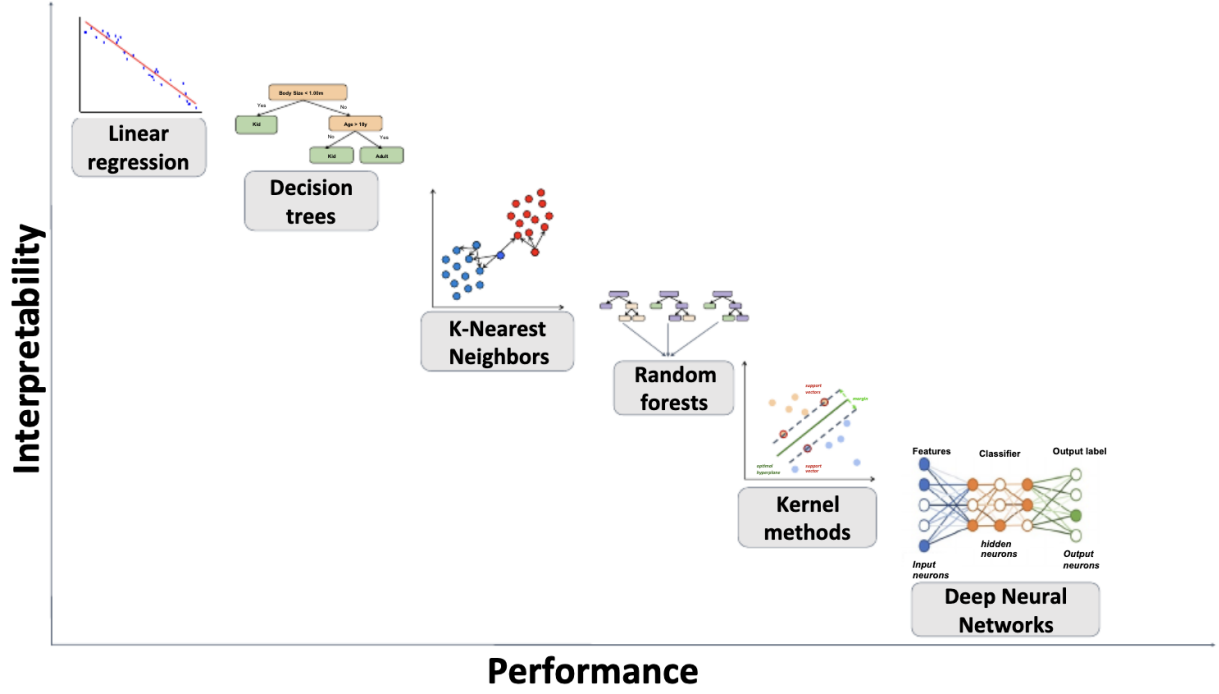


Figure 2.2: Plot showing the performance vs. Interpretability of each machine learning algorithm [6]

From the list of different algorithms above, linear regression is the simplest and most straightforward approach for predicting the relationship between the dependent variable (Y) and one or more independent variables (X), which can be shown below in equation 2.11 [7]:

$$Y \approx \beta_0 + \beta_1 X_1 + \dots + \beta_b X_n \quad (2.11)$$

Where β_0 , β_1 , and β_n are the unknown constants representing the intercept and slope terms of the linear model, the goal of the linear regression is to find the values for the unknown constants to minimize the difference between the dependent variable and the independent variable. The advantage of the linear regression model is that it is easy to understand and implement due to the assumption that the relationship between the dependent and independent variables is linear. However, this assumption may not be true for many real-world problems, leading to lower accuracy in the result.

Next is the decision tree, where the algorithm works by recursively partitioning the data into subsets based on the input values, creating a flowchart-like tree structure. The algorithm will start as one node at the top; then, it will branch out to multiple sub-nodes by selecting the best feature and threshold. If the value of the input is below a certain threshold, the left branch is followed; otherwise, the right branch is followed, as shown in Figure 2.3. The algorithm will recursively spit the data until specific criteria are met, and each tree node presents a prediction [7]. The advantage of a decision tree is that it is easy to understand and can handle missing values well. However, it is prone to overfitting and is sensitive to changes in the data.

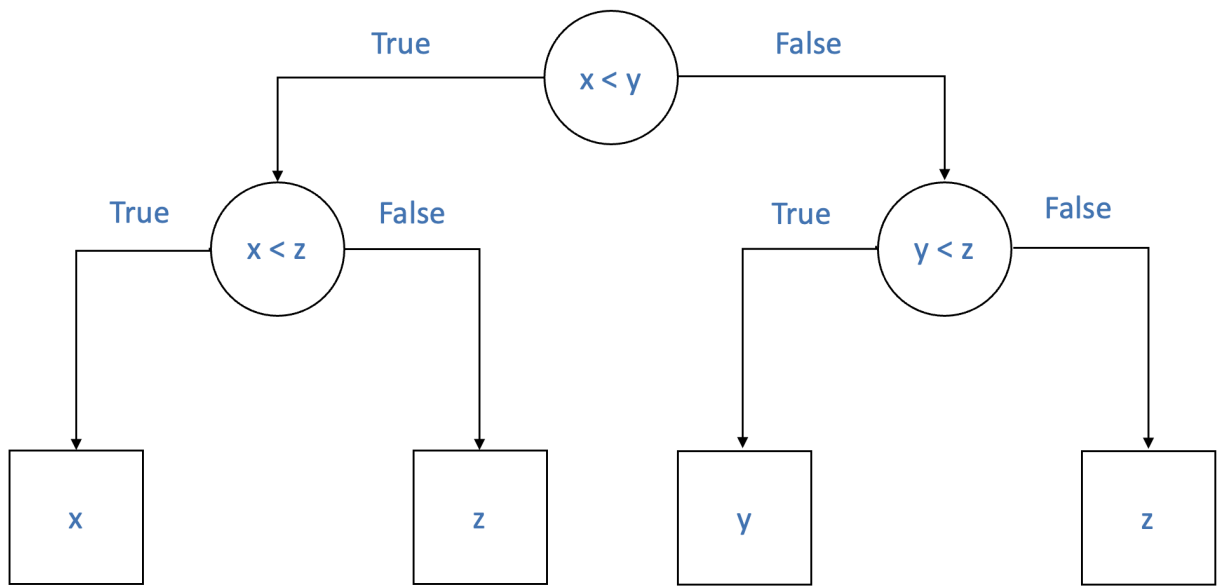


Figure 2.3: Decision Tree

The basic idea behind SVR is to find a hyperplane in p -dimensional space that will separate the data into two classes and maximizes the margin. This hyperplane can be found by subtracting one from the p -dimension. For example, in two dimensions, the hyperplane is simply a line. In Figure 2.4 [7], the solid line represents the hyperplane, while the margin is the distance from either side to the dashed lines, the blue and purple points are called support vectors. In a regression problem, the dashed lines are the margin of tolerance (ϵ), which allows the algorithm to tolerate a certain amount of errors while still maximizing the margin. The advantage of SVR

is that it will perform well if there is a clear margin in the data, effective in high dimensional space. However, the disadvantage of SVR is that it is not good for problems with a large amount of data or a lot of noise in the data set.

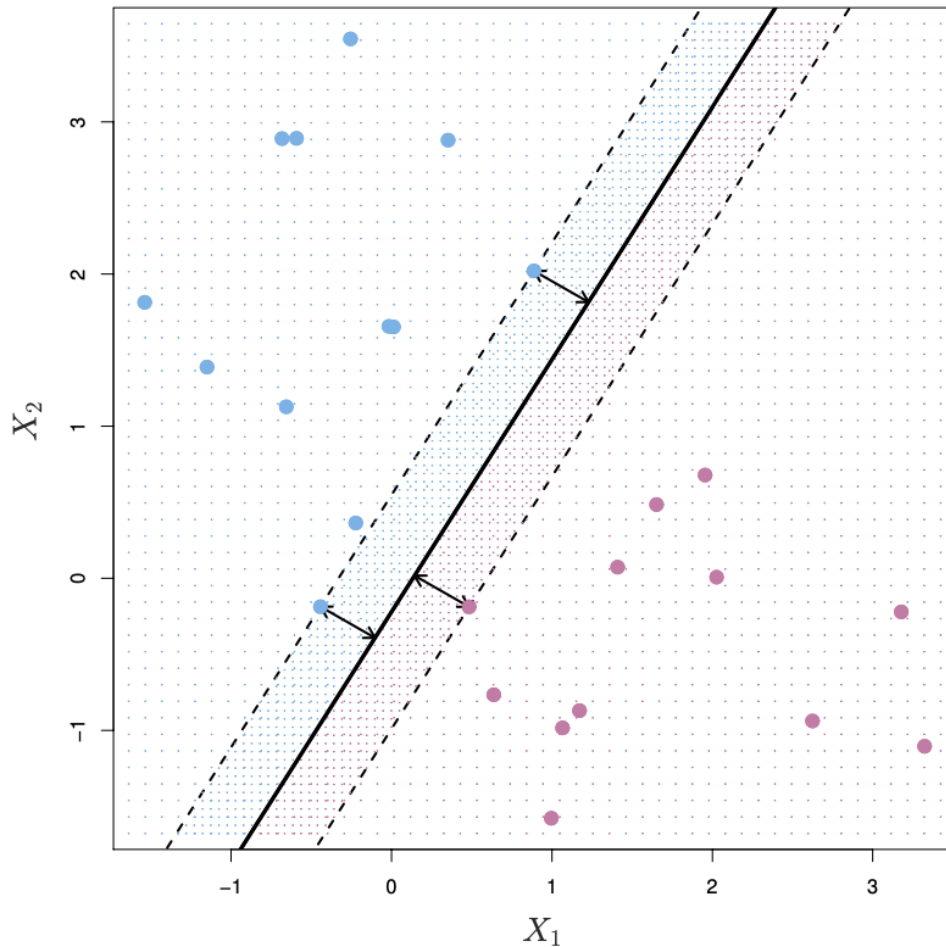


Figure 2.4: Plot explaining Support Vector Machine [7]

As the name implies, the way KNN works is the algorithm will find the k-nearest data points to a new input and use the average value of the k-nearest data points as the prediction. The parameter k is a hyperparameter that is set before the training. This hyperparameter determines the number of nearest data points consider. Figure 2.5 [7] has six blue and six orange circles with $K = 3$. The goal is to make a prediction of what the "x" mark is. The algorithm will identify the three points nearest to the "x" and use it to make a prediction of what "x" would be. In this case, 2/3 of the points are blue, and 1/3 of the points are orange. So the algorithm will predict

this point to be blue. The advantage of KNN is that it is simple to understand and implement and can handle non-linear relationships. However, KNN is sensitive to the chosen K value; if K is too high, it can cause overfitting.

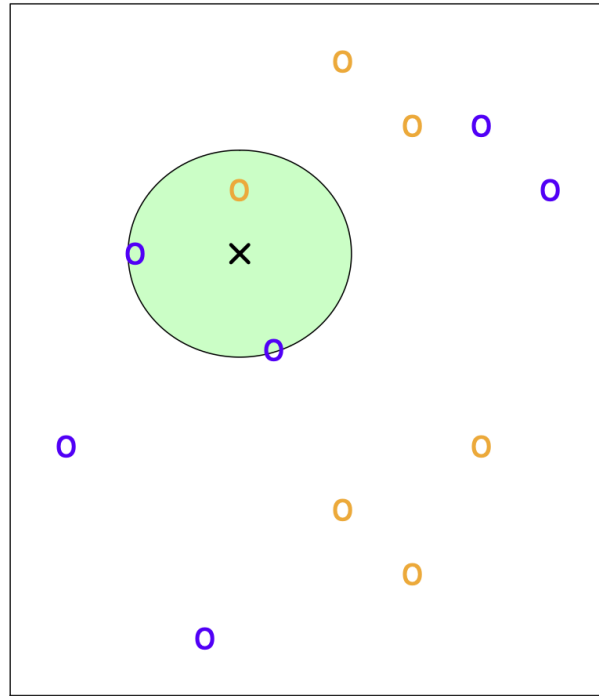


Figure 2.5: Plot explaining K-Nearest Neighbor with $K = 3$ [7]

When doing machine learning, it is essential to do data partitioning, which divides the dataset into multiple subsets for training, validating, and testing the machine learning models. The most common method is called the "validation set approach," where the data set is split into two parts, training and testing [7]. In this method, the dataset can either be split into 80/20 or 70/30, where the majority of the data set will be used for training, and the remainder will be used to see the performance of the prediction. In addition, data partitioning aims to determine if the model is overfitting, a common problem in machine learning when the model is trained too well on the training data but performs poorly on the new, unseen data. This occurs when the model is too complex or has too many parameters, and the model is able to perfectly fit all the noise and random fluctuations that occur in the training data but fails to generalize to new data, as seen in Figure 2.6 [6]. One way to prevent overfitting is to do regularization, which

reduces the number of parameters, or do data partitioning.

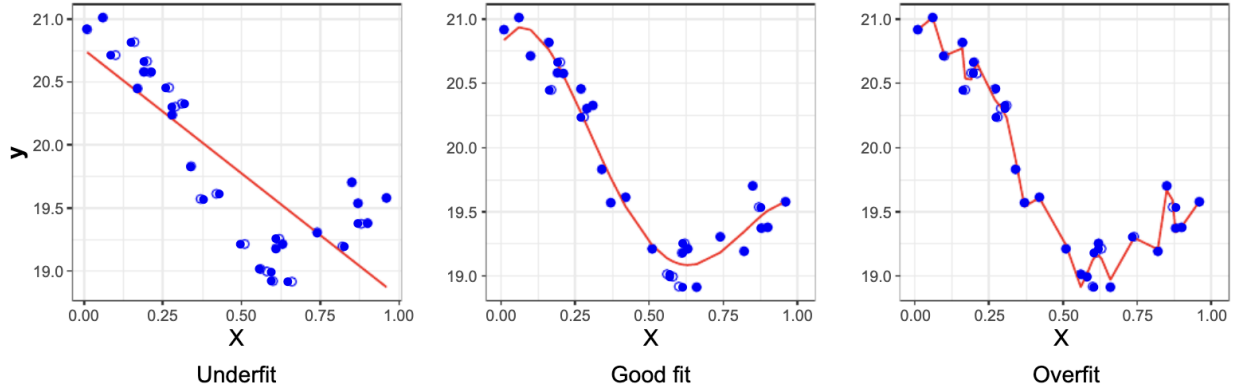


Figure 2.6: Plots demonstrating underfitting, good fit, and overfitting [6]

Once the model has been trained and makes a prediction on the unseen data, it is necessary to evaluate the performance of each algorithm. This can be done by calculating the Mean Squared Error (MSE), Root Mean Squared Error (RMSE).

The MSE measures how well the predictions actually match the observed data [6]; lower values of MSE indicate a better fit, while higher values indicate a poorer fit. However, MSE is sensitive to outliers, and a good fit model often has a high MSE if the data is highly skewed. MSE can be calculated as shown in equation 2.12 [6]:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (2.12)$$

Where n is the number of data points, $\hat{f}(x_i)$ is the prediction that \hat{f} gives for the i th observation. RMSE is similar to MSE, except the equation is squared and is often used in place of MSE because it is expressed in the same units as the target variable. It is used when the target variable has a large range of values. The equation for RMSE is shown below in equation 2.13:

$$\sqrt{MSE} = RMSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(x_i)) \quad (2.13)$$

Once the MSE or RMSE has been calculated and yielded low values, it is possible to use the model to make predictions for a different range of the five parameters and perform parameters optimization.

2.5 Optimization

Optimization is the process of finding the best solution to a problem by maximizing or minimizing an objective function. For this thesis, the objective function is to minimize the intrusion area of GDL into the GFC. This is important because the intrusion can significantly affect the PEMFC performance and degradation of the PEMFC. Using one of the four machine learning models discussed in the previous section to make predictions on a larger range of parameters will allow researchers to explore a larger range of possible parameter values and identify the combination that yields the least intrusion area. This can help researchers avoid the time-consuming process of experimenting with every possible combination of parameters. For example, during the gathering data process, it takes about fifteen minutes to create a model on COMSOL, get the intrusion area coordinates, and calculate the % of the intrusion area. With the trained model, it is possible to get a combination of parameters that will yield less than 10% or 20% of the intrusion area. Figure 2.7 shows the workflow process for this experiment.

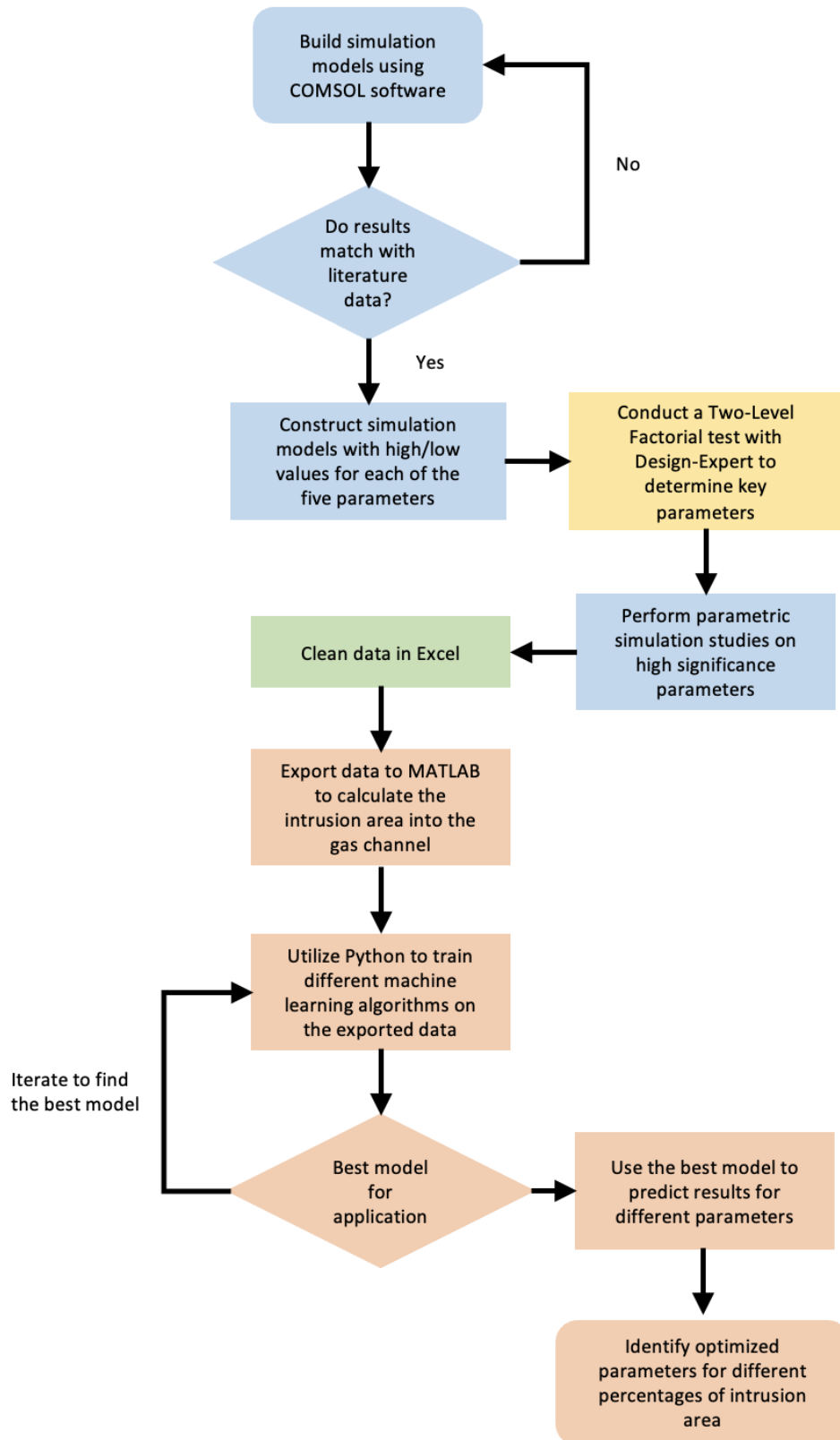


Figure 2.7: Schematic diagram of the steps for this experiment

Chapter 3

Results and Discussion

3.1 Simulations and Two-Factorial Results

To ensure the model is accurate before running parametric studies to obtain data for the machine learning portion, it is important to verify the model with literature data. Utilize the same parameters as seen in Table 3.1 from the study by Wang et al. [4] to generate the model as seen in Figure 3.1. The top portion represents the BP, and the bottom portion presents the GDL. A mesh size of 2×10^{-5} m, as seen in Figure 3.2, is employed to begin the static study. The von Mises stress result of the static study can be seen in Figure 3.3. As evident, the GDL thickness portion under the R_{BP} decreases, resulting in the intrusion of the GDL into the channel.

Table 3.1: Inputs for two-factorial design

Parameters	Value
H_{GDL}	$0.3 \times 10^{-3} m$
E_{GDL}	6.3 MPa
CW_{BP}	$1.0 \times 10^{-3} m$
R_{BP}	$0.5 \times 10^{-3} m$
P	0, 1, 2, 3 MPa

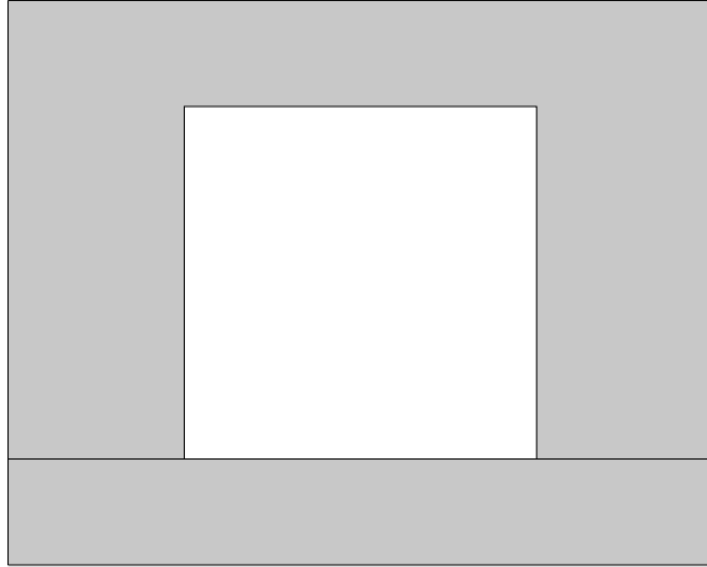


Figure 3.1: BP and GDL model from COMSOL software

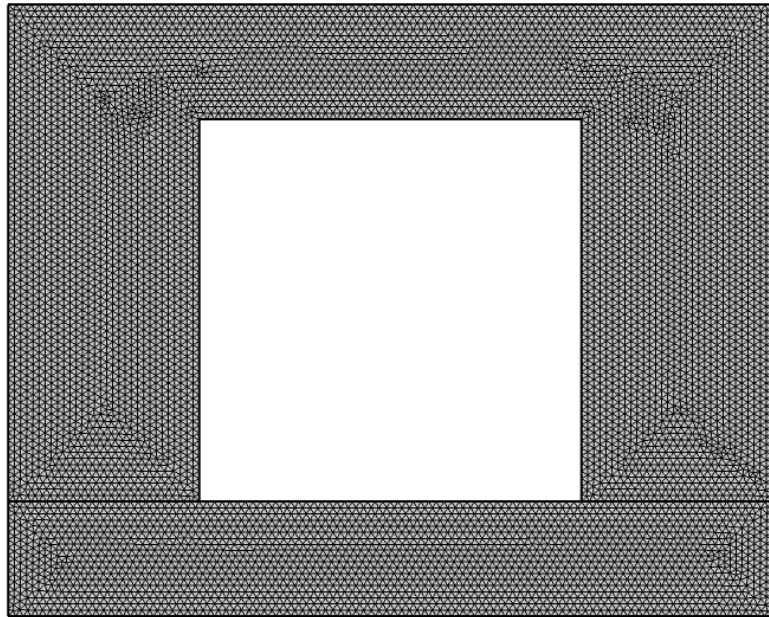


Figure 3.2: Mesh size of 2×10^{-5} m for the model

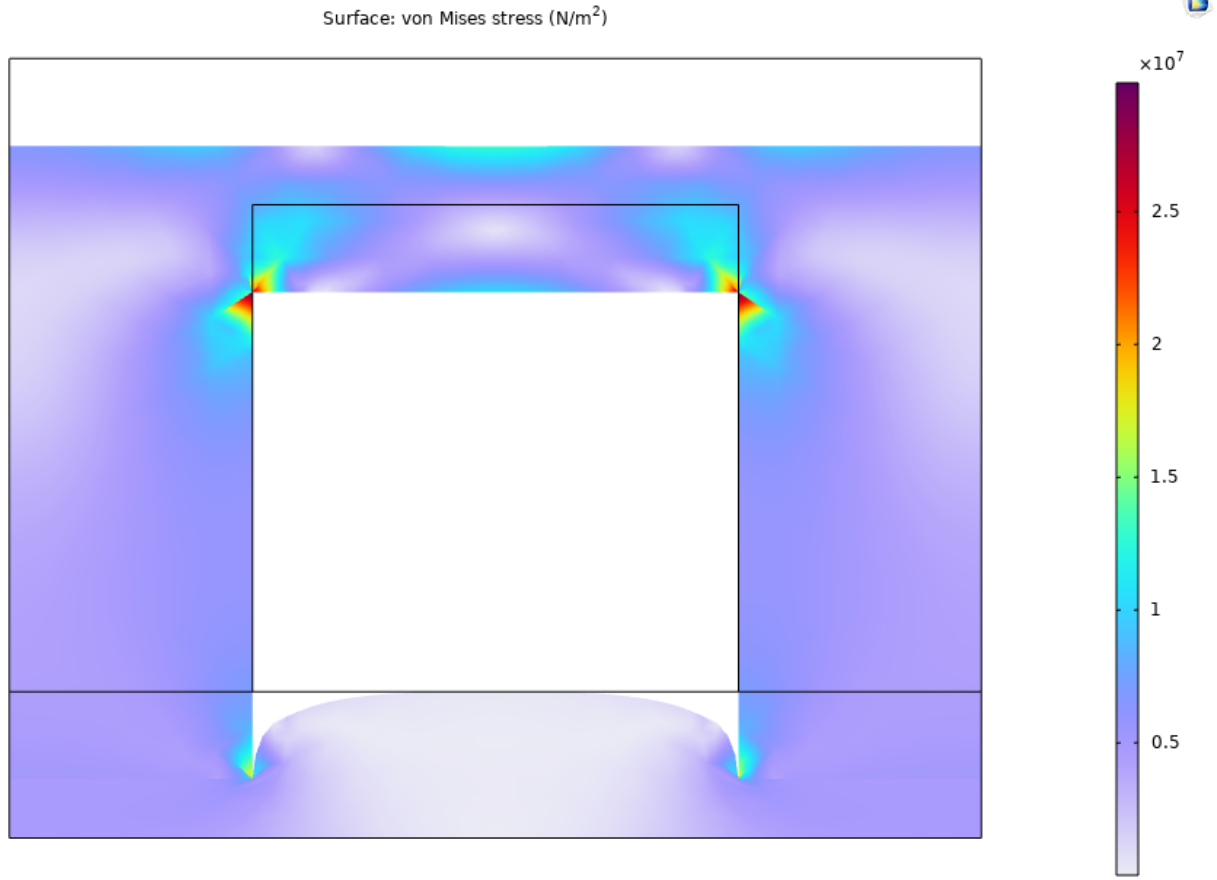


Figure 3.3: Schematic of BP and GDL

Once the simulation is completed, the x and y coordinates of each node of the intrusion arc length are used to verify the data in Wang et al. study [4]. As seen in Figure 3.4, the simulation intrusion results presented by the bold lines closely match with the literature data, represented by the dotted lines for all four different pressures. This correlation confirms the accuracy of the model and can be used for the parametric study.

After verifying the model results with literature data, perform parametric studies for the high and low values as seen in Table 2.2 to determine the key parameters that affect the intrusion area. Export the x and y coordinates of each node of the intrusion arc length from all 32 combinations of the five parameters into MATLAB. Since the intrusion area is symmetrical, only half of the arc length is used, starting from the end of the channel width to the middle of the intrusion arc length. This data is exported into MATLAB, where the *trapz* function is utilized to

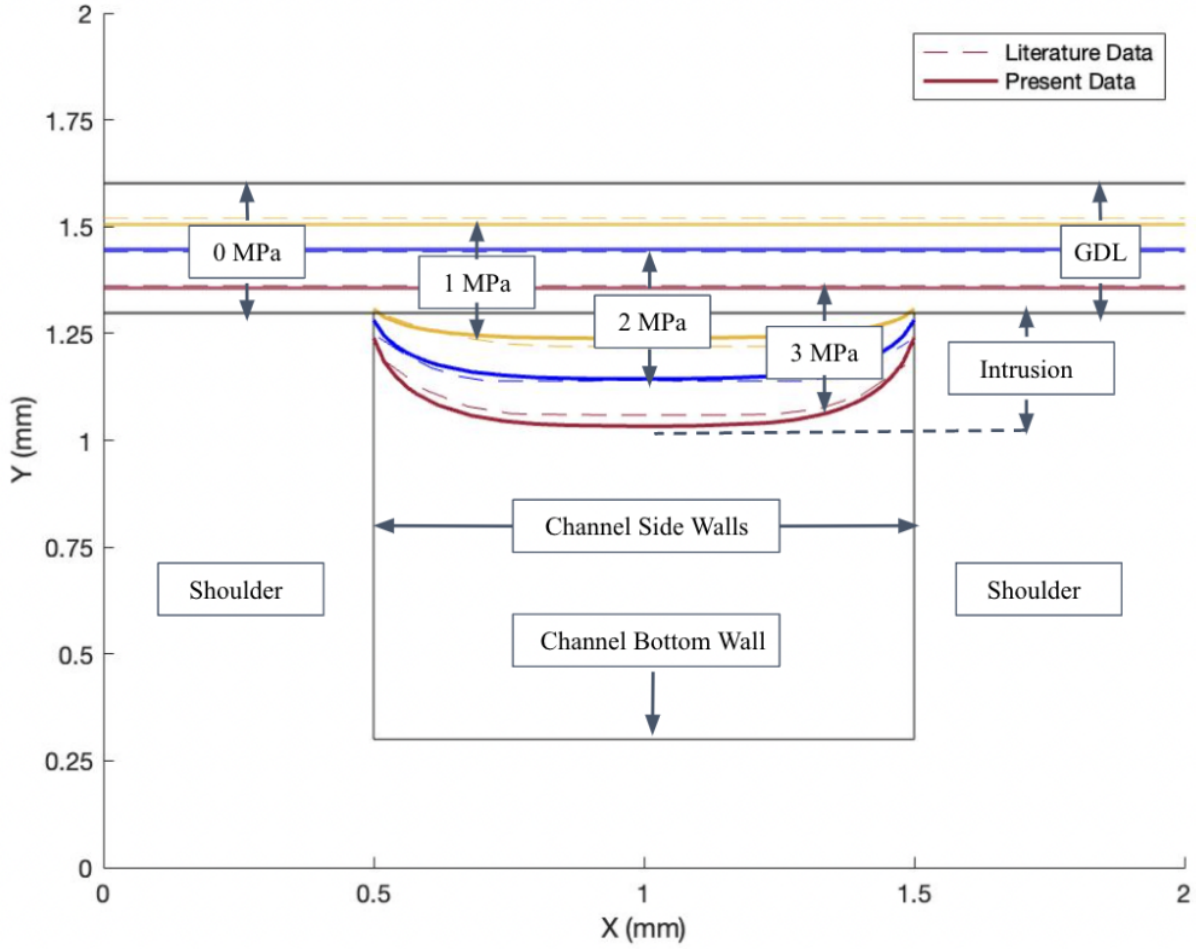


Figure 3.4: Comparison of simulation results with literature data

calculate the intrusion area. As seen in Figure 3.5, the upside-down half intrusion area of the GDL is illustrated, with the intrusion area located in section B of the plot and the dotted rectangle representing the thickness of the GDL. To find the intrusion area, the area of the rectangle is calculated and then subtracted from the area A calculated by the *trapz* function, which yields the half intrusion area which is area B. This value is then multiplied by two to obtain the entire intrusion area. Then divide this intrusion area by the area of the channel to obtain the % intrusion area into the gas channel.

Together with the five parameters, the % intrusion areas are exported into Design-Expert to determine key parameters using two-factorial. The results of the two factorials are shown in Table 3.2. The P-values for all the parameters are < 0.0001 , which is less than 0.05; this indicates that

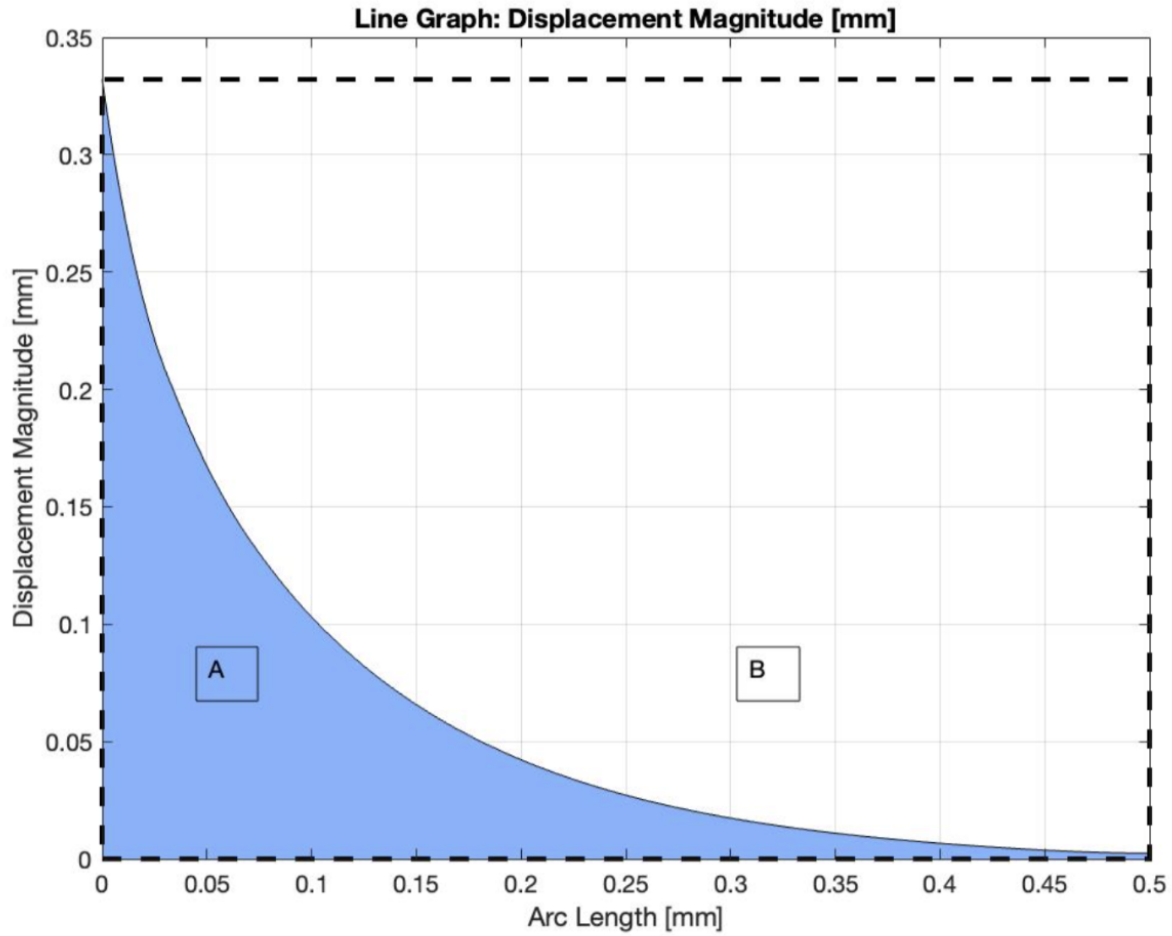


Figure 3.5: Half the intrusion area of the GDL into the channel

all the parameters are important. The E_{GDL} has the highest F-value of the five parameters. This means that this parameter has the highest impact on the % intrusion area and, therefore, will have the most extensive interval for the parametric studies. The software also shows a relationship between the parameters that affect the % intrusion. However, the F-values for these relationships are much smaller than the parameter individually, meaning they do not have as much effect. For easier visualization, the F and P values for these parameters and relationships can also be presented in a half-normal plot as seen in Figure 3.6. The further away the parameter is from the red line, the higher impact that parameter has on the % intrusion area. The plot also indicates whether or not the parameter has a positive or negative effect on the result. It is worth noting that the software sees the decrease in % intrusion as a negative effect, so it labeled

the E_{GDL} as a negative effect. However, increasing the E_{GDL} should yield a lower % intrusion area. So, in this case, the blue color represents increasing the value will lower the % intrusion area, whereas the orange color will yield a larger % intrusion area.

Table 3.2: F-values and P-values obtained from Design-Expert software

Source	F-value % intrusion	P-value % intrusion
Model	694.16	< 0.0001
A - H_{GDL}	159.96	< 0.0001
B - CW_{BP}	992.59	< 0.0001
C - R_{BP}	737.82	< 0.0001
D - E_{GDL}	3776.66	< 0.0001
E - P	387.45	< 0.0001
AB	34.17	< 0.0001
AC	46.82	< 0.0001
BC	74.38	< 0.0001
ABC	37.57	< 0.0001

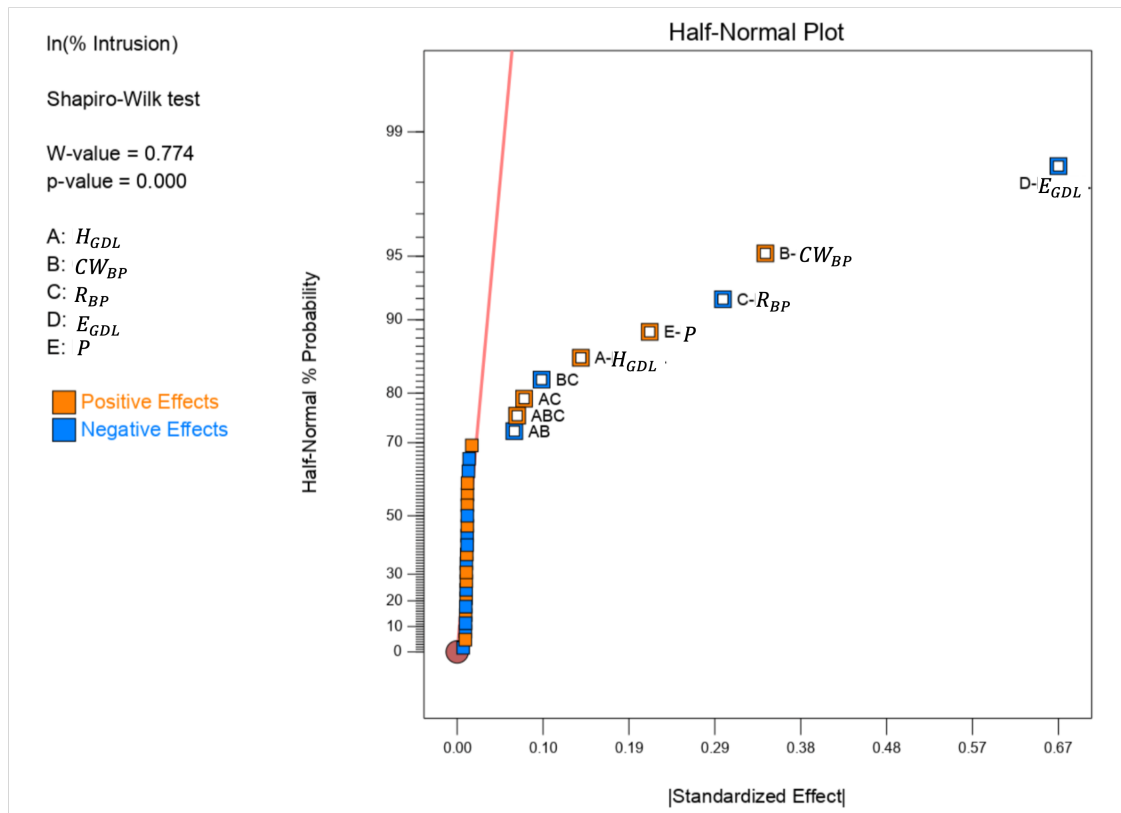


Figure 3.6: F and P values of the five parameters and relationship represented in a half-normal plot

Since the P-values obtained from the two-factorials test indicate that all five parameters affected the % intrusion area, all five will be used for the parametric study. Then the interval of each parameter is chosen based on the F-value, with the parameter with the highest F-value having the widest interval to make sure the machine learning output more accurate results. For example, in this case, the E_{GDL} will have six different inputs between 6.3 MPa to 31.5 MPa. The other parameter that has a lower F-value will have smaller intervals. The range of these intervals is chosen based on Bo et al. study [50]. In the interest of the applied pressure, it will also have a high interval despite having a lower F-value compared to CW_{BP} . The parametric study can be conducted by running simulations for each combination of parameter values, as shown in Table 3.3. The results will be used to train the machine learning models and make predictions on the % intrusion area.

Table 3.3: Inputs for the parametric study

Parameters	Value
H_{GDL} [mm]	0.2, 0.4, 0.6
E [MPa]	6.3, 11.34, 16.36, 21.38, 26.4, 31.5
CW_{BP} [mm]	1.0, 1.3, 1.5
R_{BP} [mm]	0.5, 0.75, 1.0
P [MPa]	0.5, 1, 2, 3, 4, 5

3.2 Results of Different Machine Learning Algorithms

The parametric study of the five different parameters will yield 972 different combinations of training inputs for machine learning. These data will be exported to Excel for cleaning to ensure that the x and y values correspond correctly to the specific parameter. The data will then be exported to Matlab for the intrusion area calculation before being trained by different machine learning algorithms in Python. Figure 3.7 shows the impact of each parameter on the intrusion area. As already seen in the previous section with the half-normal plot. The E_{GDL} and R_{BP} pos-

itively affect the % intrusion area, meaning as these two parameters increased, the % intrusion area decreased. Whereas the remaining three parameters will have negative effects on the % intrusion area, when they increase, the % intrusion area also increases.

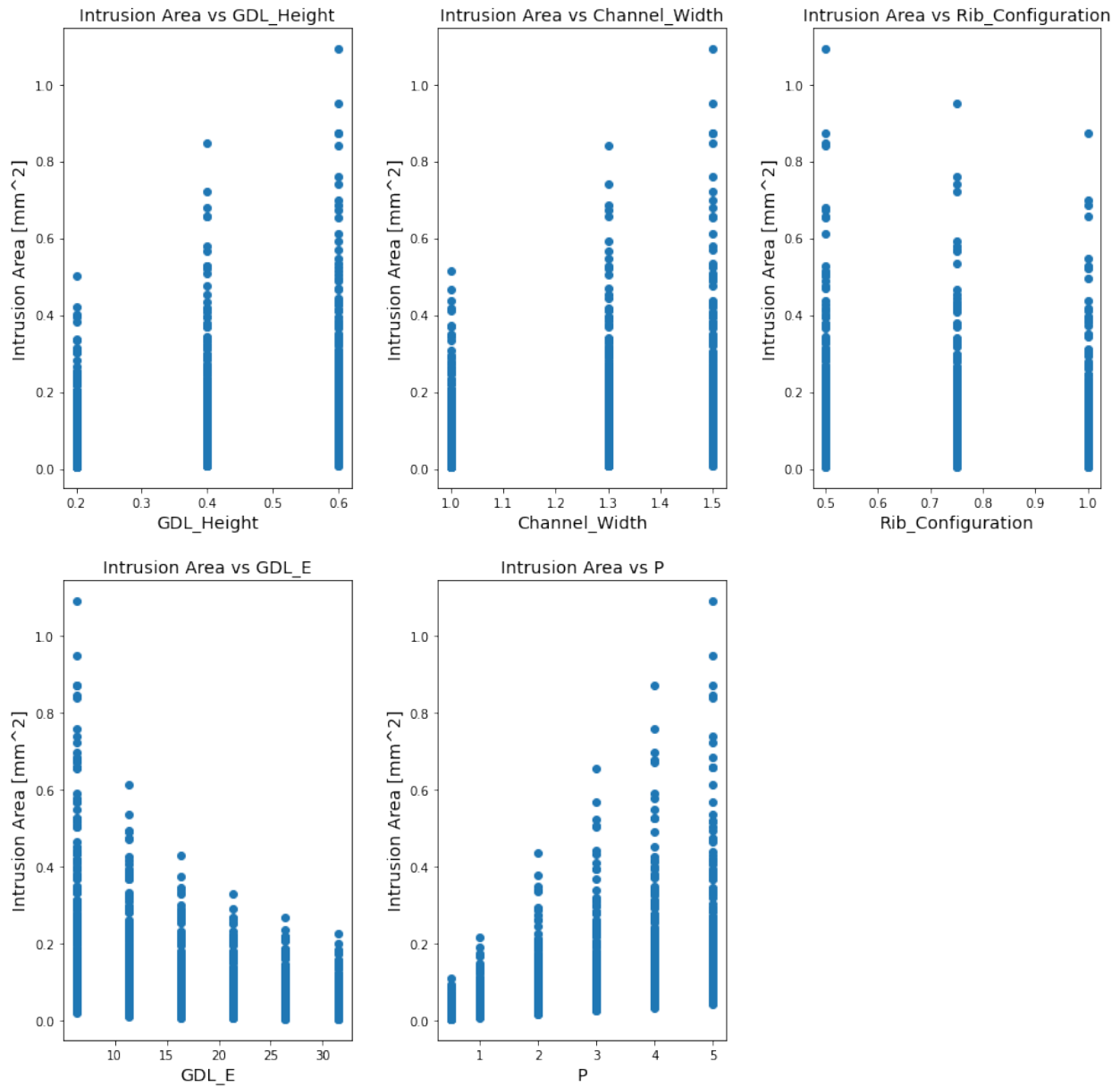


Figure 3.7: Effects of different parameters on the intrusion area

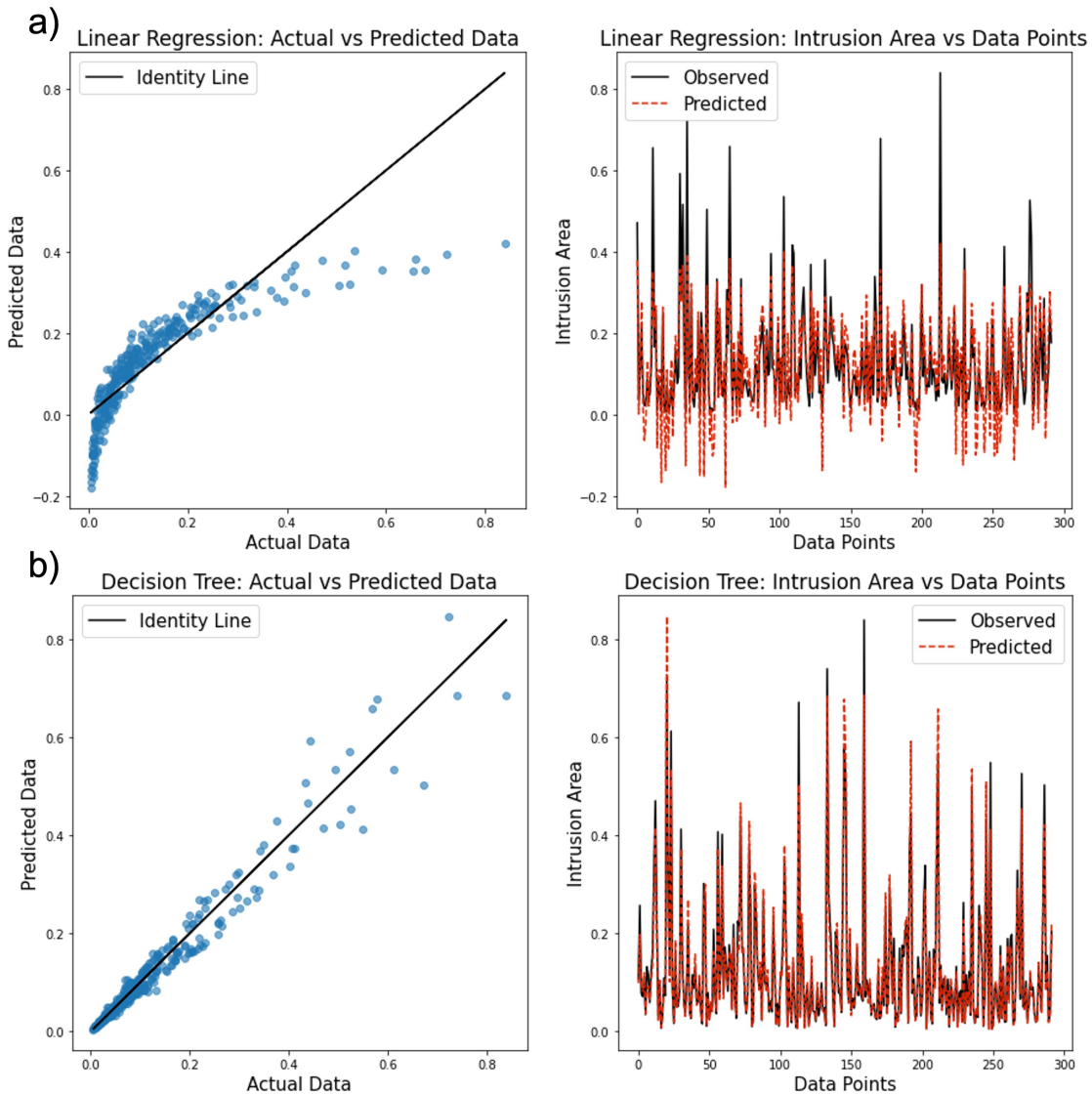
Using different *sklearn* library in Python to implement the four machine learning algorithms and obtain the accuracy of each model. Each algorithm will be trained using the same dataset, and their accuracy will be measured using the same metric. The linear regression algorithm

will be implemented using the `LinearRegression()` function, decision tree algorithm will be implemented using the `DecisionTreeRegressor()` function, random forest algorithm will be implemented using the `RandomForestRegressor()` function, and support vector regression algorithm will be implemented using the `SVR()` function. Then use the equation from chapter 2 to calculate the RMSE of each algorithm.

The models will use 70% of the data for training and the remaining 30% for testing. Figure 3.8 shows the results of the four models. The left plot represents the relationship between the actual data and the predicted data from each model. The x-axis represents the actual data or the test data, and the y-axis represents the predicted data from the model. The black diagonal line is called the identity line, and it represents the model being 100% accurate. The points on the plot show the correlation between the actual and predicted values. So the closer it is to the identity line, the closer the predicted point to the actual value. The right plot represents the comparison of the actual data and the predicted data. The x-axis represents the data points, and the y-axis represents the intrusion area. The black line represents the actual data, and the red line represents the predicted data. The plots show the correlation between the actual and predicted values. The overlap between the two lines represents the perfect correlation between the two values. The accuracy and RMSE of each model are shown in Table 3.4.

As seen in Figure 3.8, the Linear Regression model has the worst prediction out of the four models. The model prediction deviates from the identity line at around 0.3 on the x-axis. This shows that the model is not linear. This can also be observed in the Intrusion Area vs. Data Points plot, where the model inaccurately predicts negative intrusion areas at certain parameters, and there is a significant mismatch between the observed and predicted values. Additionally, as seen in Table 3.4, the Linear Regression model has the lowest accuracy and highest RMSE. The linear assumption in the Linear Regression model did not capture the complexities and non-linearity in the data. In contrast, the other three models were able to handle the complexities in the data, and their predictions from these models stay relatively linear to the identity lines but start to

deviate a little as the x-axis increases. In the Intrusion Area vs. Data Points plots, the observed and predicted values have a higher overlap compared to the Linear Regression model. From Table 3.4, it is clear to conclude that out of the four models, Decision Tree will yield the highest accuracy and lowest RMSE compared to the other three models. Using this information, Decision Tree will be used to optimize the parameters to yield the lowest intrusion area.



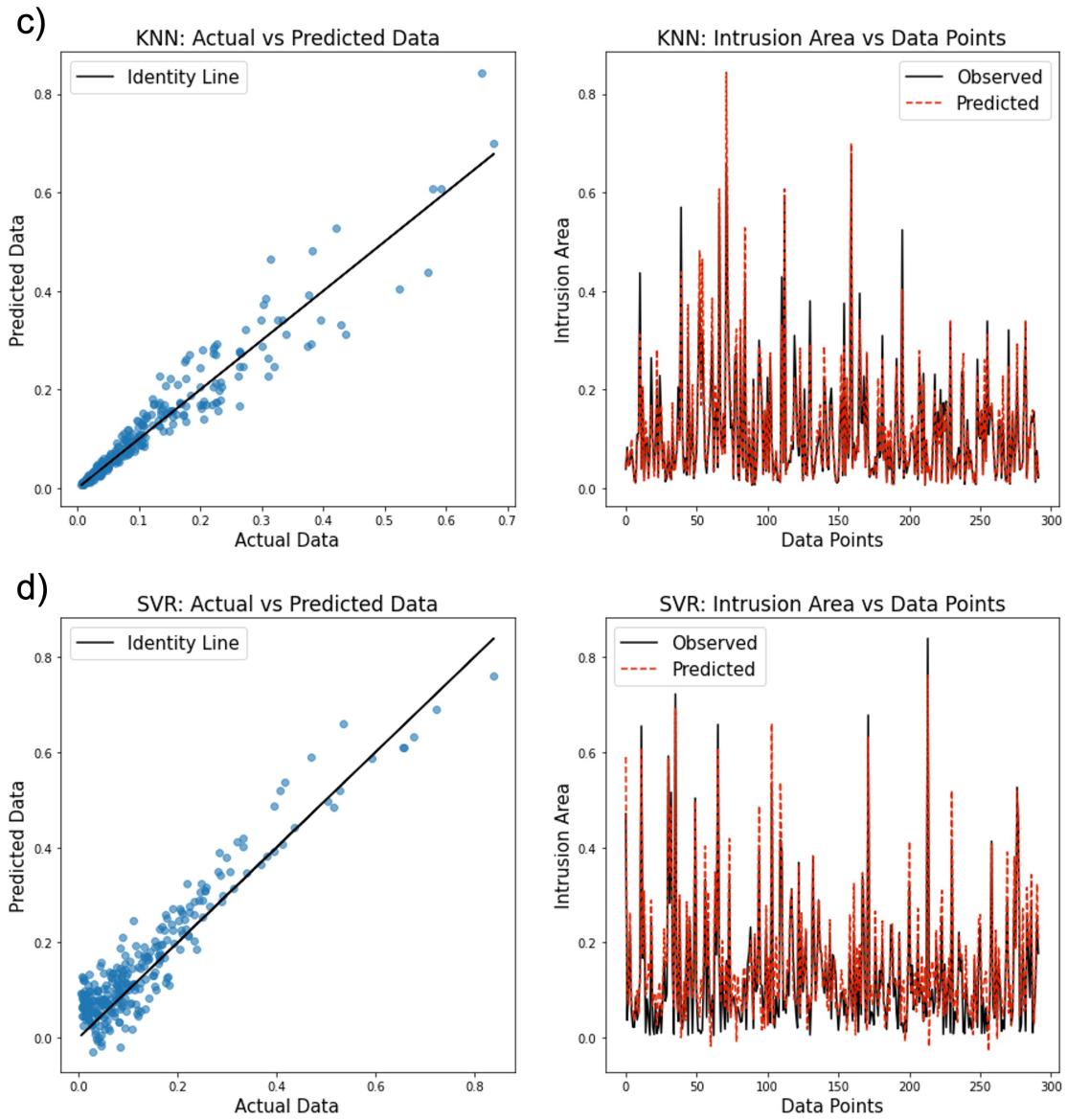


Figure 3.8: Results of each model, a) Linear Regression, b) Decision Tree, c) KNN, d) SVR

Table 3.4: Models accuracy and RMSE

Model	Model Accuracy	RMSE
Linear Regression	68.5%	0.0751
Decision Tree	95.5%	0.0303
KNN	92.9%	0.0372
SVR	85.2%	0.0516

3.3 Optimization Results

Having shown that the Decision Tree model is best, it is used to make predictions for various different parameters to figure out the combinations that will yield intrusion area that is lower than 10%, and 20%. The model will make predictions on a range of values for two parameters while keeping the other three constant. The model will make predictions on a total of 100 different combinations of each parameter then it will go through a for loop to separate the parameters that will yield either 10% or 20% intrusion area.

To utilize the plots in Figure 3.9 and 3.10, three of the parameters are set constant while varying the other two, as shown in the legend labeled Model Parameters. Then from there, pick any value from the varying parameters for % intrusion area in interest. For example, in the graph in Figure 3.9.a, the CW_{BP} , R_{BP} , and E_{GDL} are fixed at 1.5 mm, 1.0 mm, and 15 MPa respectively. When seeking intrusion area less than 10%, the H_{GDL} can be between 0 - 0.6 mm, and the P can be between 0 - 6 MPa.

Figure 3.9.b demonstrate the correlation between the parameters of R_{BP} , E_{GDL} , with the % intrusion area. As previously discussed, both R_{BP} and E_{GDL} have a positive effect on reducing the % intrusion area, with higher values of these parameters resulting in lower % intrusion. As seen in this figure, anything before $R_{BP} = 0.8\text{mm}$, and $E_{GDL} = 14\text{ MPa}$ will result in % intrusion area greater than 20%. Whereas, anything past $R_{BP} = 0.9\text{mm}$, and $E_{GDL} = 16\text{ MPa}$ will be less than 10%. The remaining plots in Figure 3.10, depict the optimization for P , CW_{BP} , and H_{GDL} . Unlike R_{BP} and E_{GDL} , these parameters have a negative effect on the % intrusion area, with higher values resulting in higher % intrusion.

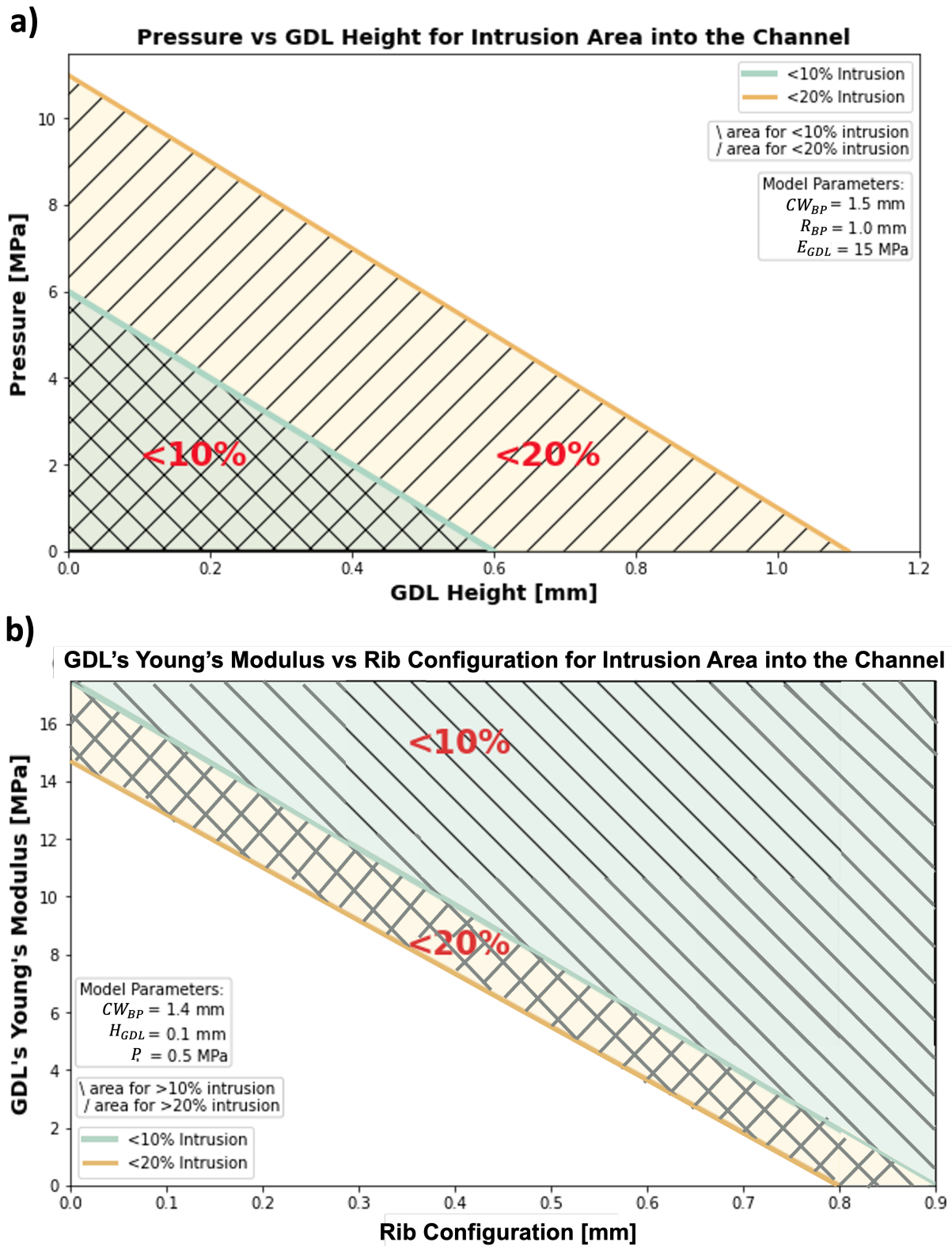


Figure 3.9: Optimization results for a) Pressure vs GDL Height and b) GDL Young's Modulus vs Rib Configuration

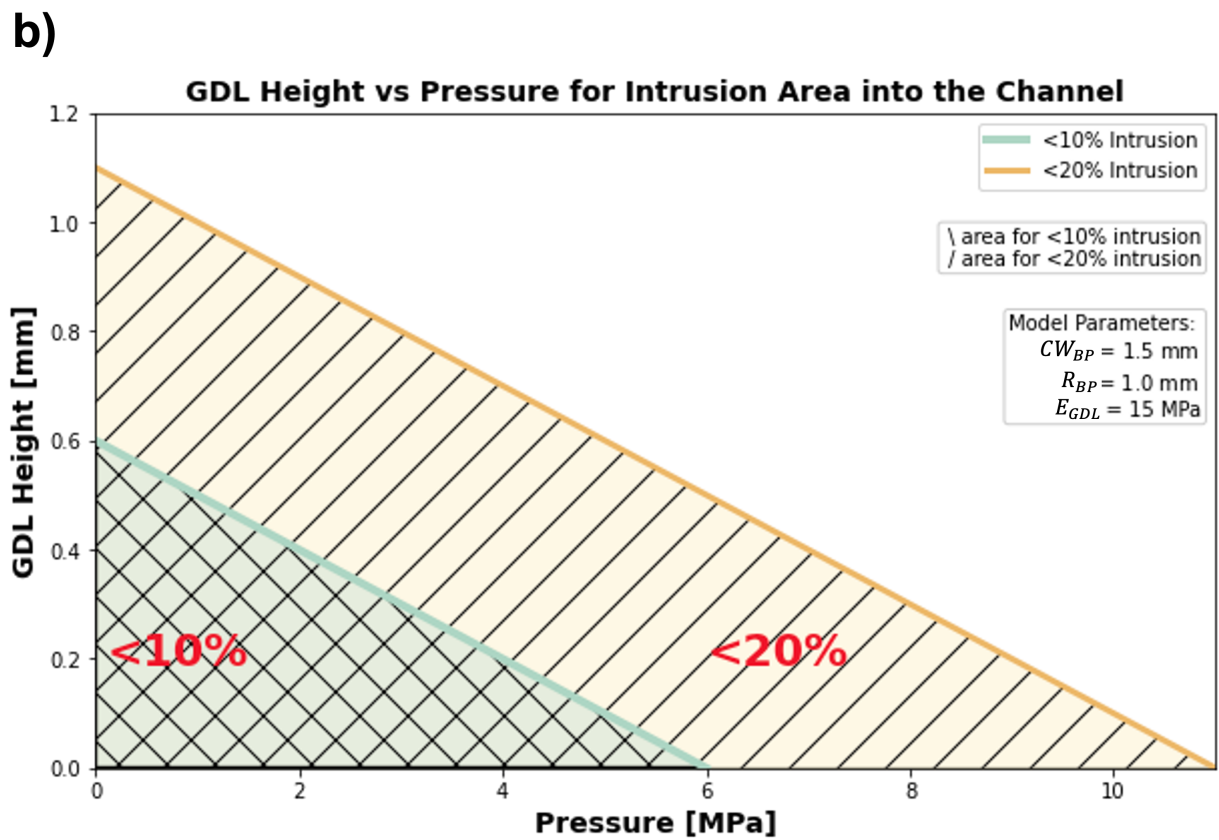
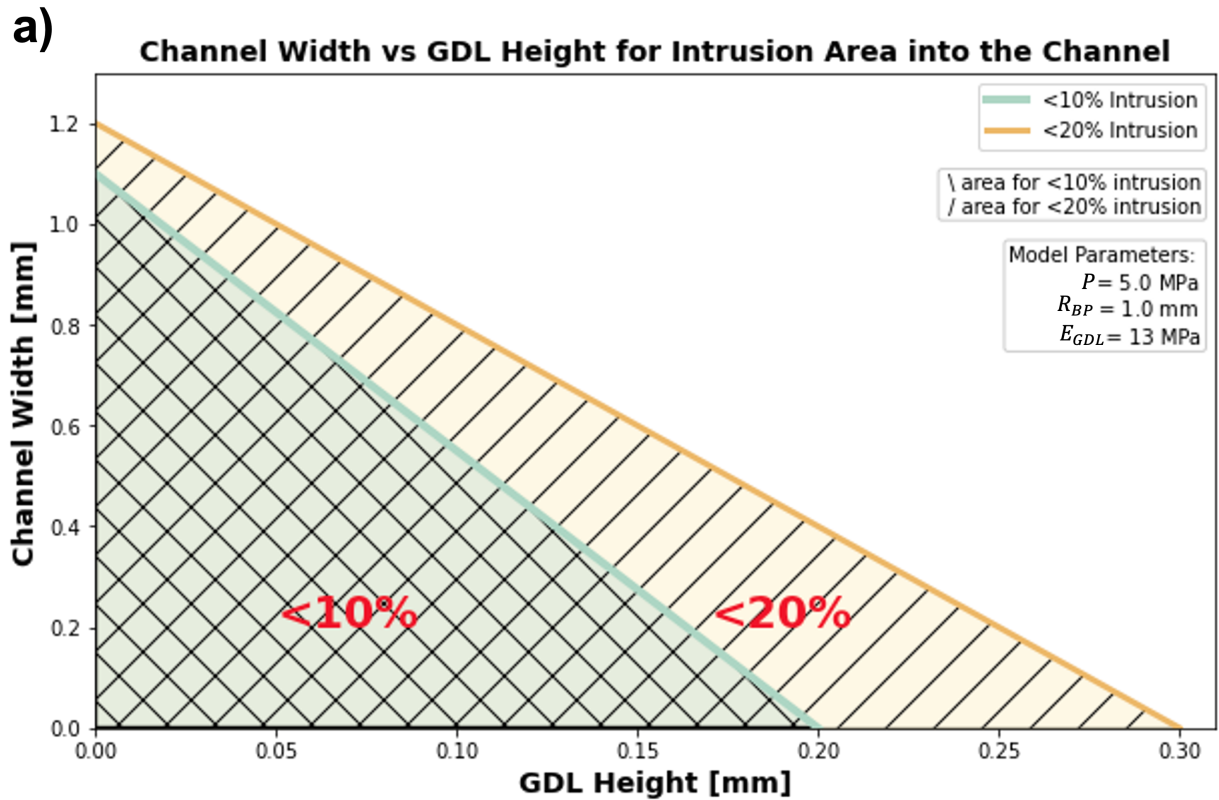


Figure 3.10: Optimization results for a) Channel Width vs GDL Height and b) GDL Height vs Pressure

Chapter 4

Conclusions

In this thesis, focus was placed on understanding the impact of different parameters on the Gas Diffusion Layer (GDL) intrusion area into a gas channel for PEMFC. COMSOL was used to model the physical behavior of the GDL deformation and gather data on the intrusion area. Design Expert software was used for two-factorial tests, and Python machine-learning algorithms were used to analyze the data from these simulations.

The parameters that were studied in this research were Bipolar Plate Rib (R_{BP}), GDL Young's Modulus (E_{GDL}), Pressure (P), GDL Height (H_{GDL}), and BP Channel Width (CW_{BP}). From both the two factorial tests and machine learning, it was found that among these parameters, E_{GDL} had the greatest impact on the intrusion area. This was determined by comparing each parameter's F-values, which measure how much each parameter contributes to the overall variation in the data. It was also found that both the CW_{BP} and the E_{GDL} have a positive effect on the intrusion area. This means that the intrusion area is decreased when these parameters are increased. On the other hand, the other parameters P, R_{BP} , and H_{GDL} were found to have a negative impact on the intrusion area, which means that when these parameters are increased, the intrusion area is also increased.

Using the finding from the Design Expert software to design the range of each parameter for the parametric study to train the machine learning models. Since E_{GDL} has the highest impact, it had larger range than the other parameters. Four machine-learning algorithms were employed to analyze the data further and find the optimal design parameters: linear regression, Decision Tree, SVR, and KNN. These algorithms were trained using 70% of the data set and were tested using the remaining 30%. The results of this study found that the Decision Tree algorithm had the highest model accuracy at 95.5% and the lowest Root Mean Squared Error (RMSE) at 0.0303. In contrast, linear regression had the lowest model accuracy at 68.5% and the highest RMSE at 0.0751. The Decision Tree algorithm made predictions for various ranges of the five parameters to find the optimal design parameters. This study was done to reduce the time it takes to simulate the model and find the intrusion area. Typically, it takes an average of 20 minutes to create the model, simulate it, and calculate the intrusion area. However, with the trained models, the intrusion area can be found in less than a minute. This significantly improves the efficiency and accuracy of the design process.

Overall, this research provides valuable insights into how different parameters affect the intrusion area in a PEMFC. By understanding the impact of these parameters, researchers and engineers can design PEMFCs that are more efficient from having a lower intrusion area. The use of machine-learning algorithms in this research has proven to be suitable for predicting the intrusion area of the gas channels in a PEMFC. It also shows the potential for machine learning to improve the design process and increase the efficiency of PEMFC research. Additionally, through two-factorial tests and four different machine-learning algorithms, optimal design parameters were determined, and predictions were made for various ranges of the five parameters studied. This research has provided valuable insights into the behavior of PEMFCs and shows the benefits of using machine learning in PEMFC research.

References

- [1] Yun Wang, Suman Basu, and Chao-Yang Wang. Modeling two-phase flow in pem fuel cell channels. *Journal of Power Sources*, 179(2):603–617, 2008.
- [2] Yun Wang, Chao-Yang Wang, and KS Chen. Elucidating differences between carbon paper and carbon cloth in polymer electrolyte fuel cells. *Electrochimica Acta*, 52(12):3965–3975, 2007.
- [3] Diogo M Neto, Marta C Oliveira, José L Alves, and Luís F Menezes. Numerical study on the formability of metallic bipolar plates for proton exchange membrane (pem) fuel cells. *Metals*, 9(7):810, 2019.
- [4] Jiatang Wang, Jinliang Yuan, Jong-Sung Yu, and Bengt Sundén. Investigation of effects of non-homogenous deformation of gas diffusion layer in a pem fuel cell. *International Journal of Energy Research*, 41(14):2121–2137, 2017.
- [5] SG Kandlikar, Z Lu, TY Lin, D Cooke, and M Daino. Uneven gas diffusion layer intrusion in gas channel arrays of proton exchange membrane fuel cell and its effects on flow distribution. *Journal of Power Sources*, 194(1):328–337, 2009.
- [6] Yalin Baştanlar and Mustafa Özuysal. Introduction to machine learning. *miRNomics: MicroRNA biology and computational analysis*, pages 105–128, 2014.
- [7] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.

- [8] Antonella Rossati. Global warming and its health impact. *The international journal of occupational and environmental medicine*, 8(1):7, 2017.
- [9] Benjamin M Sanderson and Brian C O'Neill. Assessing the costs of historical inaction on climate change. *Scientific reports*, 10(1):1–12, 2020.
- [10] AG Olabi. Energy quadrilemma and the future of renewable energy, 2016.
- [11] Ahmad Baroutaji, Arun Arjunan, John Robinson, Tabbi Wilberforce, Mohammad Ali Abdelkareem, and Abdul Ghani Olabi. Pemfc poly-generation systems: Developments, merits, and challenges. *Sustainability*, 13(21):11696, 2021.
- [12] Saad Mekhilef, Rahman Saidur, and Azadeh Safari. Comparative study of different fuel cell technologies. *Renewable and Sustainable Energy Reviews*, 16(1):981–989, 2012.
- [13] H-J Neef. International overview of hydrogen and fuel cell research. *Energy*, 34(3):327–333, 2009.
- [14] JJ Baschuk and Xianguo Li. A general formulation for a mathematical pem fuel cell model. *Journal of power sources*, 142(1-2):134–153, 2005.
- [15] Trung Van Nguyen and Mack W Knobbe. A liquid water management strategy for pem fuel cell stacks. *Journal of Power Sources*, 114(1):70–79, 2003.
- [16] Jean-Marc Le Canut, Rami M Abouatallah, and David A Harrington. Detection of membrane drying, fuel cell flooding, and anode catalyst poisoning on pemfc stacks by electrochemical impedance spectroscopy. *Journal of The Electrochemical Society*, 153(5):A857, 2006.
- [17] Peng Hu, Guang-Yi Cao, Xin-Jian Zhu, and Mingruo Hu. Coolant circuit modeling and temperature fuzzy control of proton exchange membrane fuel cells. *International Journal of Hydrogen Energy*, 35(17):9110–9123, 2010.

- [18] VM Vishnyakov. Proton exchange membrane fuel cells. *Vacuum*, 80(10):1053–1065, 2006.
- [19] Allen J Bard, Larry R Faulkner, et al. Fundamentals and applications: electrochemical methods. *Electrochem. Methods*, 2(482):580–632, 2001.
- [20] Jie Song, Qing Ye, Kun Wang, Zhiyuan Guo, and Meiling Dou. Degradation investigation of electrocatalyst in proton exchange membrane fuel cell at a high energy efficiency. *Molecules*, 26(13):3932, 2021.
- [21] Frano Barbir. *PEM fuel cells: theory and practice*. Academic press, 2012.
- [22] S Jamai Peighambardoust, Soosan Rowshanzamir, and Mehdi Amjadi. Review of the proton exchange membranes for fuel cell applications. *International journal of hydrogen energy*, 35(17):9349–9384, 2010.
- [23] Yun Wang and Ken S Chen. *PEM fuel cells: thermal and water management fundamentals*. Momentum Press, 2013.
- [24] Gang Luo, Yan Ji, Chao-Yang Wang, and Puneet K Sinha. Modeling liquid water transport in gas diffusion layers by topologically equivalent pore network. *Electrochimica Acta*, 55(19):5332–5341, 2010.
- [25] Guobin Zhang, Zhiguo Qu, Wen-Quan Tao, Xueliang Wang, Lizhen Wu, Siyuan Wu, Xu Xie, Chasen Tongsh, Wenming Huo, Zhiming Bao, et al. Porous flow field for next-generation proton exchange membrane fuel cells: Materials, characterization, design, and challenges. *Chemical Reviews*, 2022.
- [26] Allen Hermann, Tapas Chaudhuri, and Priscila Spagnol. Bipolar plates for pem fuel cells: A review. *International journal of hydrogen Energy*, 30(12):1297–1302, 2005.
- [27] Meng-Tsun Lin, Chieh-Hao Wan, and Weite Wu. Comparison of corrosion behaviors between ss304 and ti substrate coated with (ti, zr) n thin films as metal bipolar plate for unitized regenerative fuel cell. *Thin solid films*, 544:162–169, 2013.

- [28] Tabbi Wilberforce, Zaki El Hassan, Emmanuel Ogungbemi, O Ijaodola, FN Khatib, A Durrant, J Thompson, A Baroutaji, and AG Olabi. A comprehensive study of the effect of bipolar plate (bp) geometry design on the performance of proton exchange membrane (pem) fuel cells. *Renewable and Sustainable Energy Reviews*, 111:236–260, 2019.
- [29] Sirivatch Shimpalee and JW Van Zee. Numerical studies on rib & channel dimension of flow-field on pemfc performance. *International Journal of Hydrogen Energy*, 32(7):842–856, 2007.
- [30] Ahmad Baroutaji, James Carton, Joseph Stokes, and Abdul-Ghani Olabi. Design and development of proton exchange membrane fuel cell using open pore cellular foam as flow plate material. *Journal of Energy Challenges and Mechanics*, 1(7), 2014.
- [31] Yun Wang, Liem Pham, Guilherme Porto Salerno de Vasconcellos, and Marc Madou. Fabrication and characterization of micro pem fuel cells using pyrolyzed carbon current collector plates. *Journal of Power Sources*, 195(15):4796–4803, 2010.
- [32] Sung Chan Cho and Yun Wang. Two-phase flow dynamics in a micro channel with heterogeneous surfaces. *International Journal of Heat and Mass Transfer*, 71:349–360, 2014.
- [33] Jingtian Wu and Yun Wang. Liquid blockage and flow maldistribution of two-phase flow in two parallel thin micro-channels. *Applied Thermal Engineering*, 182:116127, 2021.
- [34] Chi-Hui Chien, Yao-Lun Hu, Ting-Hsuan Su, Hsuan-Ting Liu, Chung-Ting Wang, Ping-Feng Yang, and Ying-Xu Lu. Effects of bolt pre-loading variations on performance of gdl in a bolted pemfc by 3-d fem analysis. *Energy*, 113:1174–1187, 2016.
- [35] Wei Sun, Brant A Peppley, and Kunal Karan. Modeling the influence of gdl and flow-field plate parameters on the reaction distribution in the pemfc cathode catalyst layer. *Journal of Power Sources*, 144(1):42–53, 2005.

- [36] A Bazylak, D Sinton, Z-S Liu, and N Djilali. Effect of compression on liquid water transport and microstructure of pemfc gas diffusion layers. *Journal of Power Sources*, 163(2):784–792, 2007.
- [37] Alex Bates, Santanu Mukherjee, Sunwook Hwang, Sang C Lee, Osung Kwon, Gyeung Ho Choi, and Sam Park. Simulation and experimental analysis of the clamping pressure distribution in a pem fuel cell stack. *International journal of hydrogen energy*, 38(15):6481–6493, 2013.
- [38] Lianhong Zhang, Ying Liu, Haimin Song, Shuxin Wang, Yuanyuan Zhou, and S Jack Hu. Estimation of contact resistance in proton exchange membrane fuel cells. *Journal of Power Sources*, 162(2):1165–1171, 2006.
- [39] W Michael Lai, David Rubin, and Erhard Kreml. *Introduction to continuum mechanics*. Butterworth-Heinemann, 2009.
- [40] Woo-kum Lee, Chien-Hsien Ho, JW Van Zee, and Mahesh Murthy. The effects of compression and gas diffusion layers on the performance of a pem fuel cell. *Journal of power sources*, 84(1):45–51, 1999.
- [41] Jiabin Ge, Andrew Higier, and Hongtan Liu. Effect of gas diffusion layer compression on pem fuel cell performance. *Journal of power sources*, 159(2):922–927, 2006.
- [42] Sylvie Escribano, Jean-François Blachot, Jérémy Ethève, Arnaud Morin, and Renaut Mosdale. Characterization of pemfcs gas diffusion layers properties. *Journal of Power Sources*, 156(1):8–13, 2006.
- [43] P Zhou and CW Wu. Numerical study on the compression effect of gas diffusion layer on pemfc performance. *Journal of Power Sources*, 170(1):93–100, 2007.
- [44] N Khajeh-Hosseini-Dalasm, MJm Kermani, D Ghadiri Moghaddam, and JM Stockie. A

- parametric study of cathode catalyst layer structural parameters on the performance of a pem fuel cell. *International journal of hydrogen energy*, 35(6):2417–2427, 2010.
- [45] Iwao Nitta, Tero Hottinen, Olli Himanen, and Mikko Mikkola. Inhomogeneous compression of pemfc gas diffusion layer: Part i. experimental. *Journal of Power Sources*, 171(1):26–36, 2007.
- [46] Peixin He, Cheng Sun, and Yun Wang. Material distortion in laser-based additive manufacturing of fuel cell component: Three-dimensional numerical analysis. *Additive Manufacturing*, 46:102188, 2021.
- [47] Imdat Taymaz and Merthan Benli. Numerical study of assembly pressure effect on the performance of proton exchange membrane fuel cell. *Energy*, 35(5):2134–2140, 2010.
- [48] Douglas C Montgomery. *Design and analysis of experiments*. John wiley & sons, 2017.
- [49] Mark J Anderson and Patrick J Whitcomb. *DOE simplified: practical tools for effective experimentation*. CRC press, 2017.
- [50] Bo Zheng, Zhe Wang, and Yun Wang. Porous media flow field for polymer electrolyte membrane fuel cell: Depression of gas diffusion layer intrusion, deformation, and delamination. *International Journal of Energy Research*, 2022.
- [51] Yun Wang, Bongjin Seo, Bowen Wang, Nada Zamel, Kui Jiao, and Xavier Cordobes Adroher. Fundamentals, materials, and machine learning of polymer electrolyte membrane fuel cell technology. *Energy and AI*, 1:100014, 2020.
- [52] Yiheng Pang, Liang Hao, and Yun Wang. Convolutional neural network analysis of radiography images for rapid water quantification in pem fuel cell. *Applied Energy*, 321:119352, 2022.
- [53] Pengjie Tian, Xuejun Liu, Kaiyao Luo, Hongkun Li, and Yun Wang. Deep learning from three-dimensional multiphysics simulation in operational optimization and control of

polymer electrolyte membrane fuel cell for maximum power. *Applied Energy*, 288:116632, 2021.

- [54] Vishvak Kannan, Hansong Xue, K Ashoke Raman, Jiasheng Chen, Adrian Fisher, and Erik Birgersson. Quantifying operating uncertainties of a pemfc–monte carlo-machine learning based approach. *Renewable Energy*, 158:343–359, 2020.
- [55] Arun Saco, P Shanmuga Sundari, and Anand Paul. An optimized data analysis on a real-time application of pem fuel cell design by using machine learning algorithms. *Algorithms*, 15(10):346, 2022.