

IMPLEMENTATION OF AI IN SMOKE POINT PREDICTION

BY

RUAA ALA ELDIN AGEEB ABAKAR

A Thesis Presented to the
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE
CHEMICAL ENGINEERING

December 2024

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

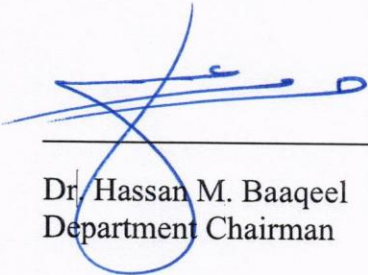
DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Ruaa Ala Eldin Ageeb Abakar** under the direction of her thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN CHEMICAL ENGINEERING**.



Dr. Abdul Gani Abdul Jameel
(Advisor)



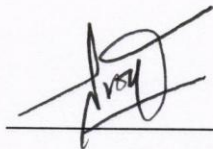
Dr. Hassan M. Baaqeel
Department Chairman



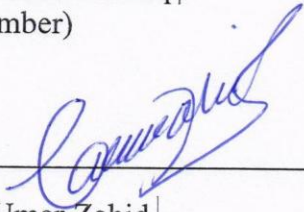
Dr. Suliman Al-Homidan
Dean of Graduate Studies

December 18th, 2024

Date



Dr. Wasif Farooq
(Member)



Dr. Umer Zahid
(Member)



© RUAA ALA ELDIN AGEEB ABAKAR
2024

Dedication

To the soul of my father, Ala Eldin, and my grandfather, Mahgoub, whose kindness, wisdom, and values continue to guide me even in their absence.

To my mother Sanaa, the angel of life whose essence of love, kindness, prayers, dedication, and invaluable support happened to be a great pillar and a key to my success. Also, upon my downtime, her smile brightens my day and heals my wound. To the dearest of all loved ones.

To my sister Razaan and brothers Mohamed and Mohannad, whose love and unwavering support fueled my passion and strengthened me throughout this journey.

To myself, for my determination and resilience, which carried me through to the completion of this thesis.

ACKNOWLEDGMENTS

All praise is gratefully and humbly given to Allah, the Omnipotent and the Sustainer. May His peace and blessings be upon the Prophet Mohammad, his family, and his companions.

My deepest thanks go to my thesis supervisor, Dr. Abdul Gani Abdul Jameel, whose expertise, patience, and insightful guidance have been invaluable throughout this journey.

I am also deeply grateful to my committee members, Dr. Umer Zahid and Dr. Wasif Farooq, and my academic advisors, Professor Mohammad Mozahar Hossain and Dr. Nayef M. Al-Saifi, for their valuable suggestions and constructive feedback. A special note of appreciation goes to Syed Amanullah, the lab technician, for his invaluable assistance, and readiness to help, which ensured smooth operations in the lab. I also wish to thank our secretary, Hasnuddin M. H. Khan, for his cooperation and efficiency, which have greatly eased the administrative aspects of my work. Moreover, I am profoundly grateful to King Fahd University of Petroleum and Minerals, for funding my master's degree. Financial support and access to world-class resources have been significant in enabling me to pursue and complete this thesis. I would also like to extend my heartfelt gratitude to the prestigious research centers, SDAIA Joint Research Center for AI and The Renewable Energy Technical Incubator (RETI), funded by the National Industrial Development and Logistics Program (NIDLP) under the Interdisciplinary Research Center for Sustainable Energy Systems (IRC-SES), for their invaluable support and the exceptional facilities they

provided during this research. Lastly, I am grateful to my colleagues and friends, whose encouragement and discussions have made this journey enjoyable. Their support has been a constant source of motivation.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	V
TABLE OF CONTENTS	VI
LIST OF TABLES	IX
LIST OF FIGURES	X
LIST OF ABBREVIATIONS	XI
ABSTRACT	XIII
ملخص الرسالة	XV
1 CHAPTER 1: INTRODUCTION.....	1
1.1 Background	1
1.2 Problem Statement.....	5
1.3 Aims and Objectives	6
2 CHAPTER 2: LITERATURE REVIEW	7
2.1 Soot Formation.....	8
2.2 Combustion Properties: Smoke Point (SP).....	12
2.2.1 SP as a measure of Sooting Index.....	12
2.2.2 Methods to Improve SP	20
2.2.3 Influence of Functional Groups on Combustion Properties	22
2.2.4 Functional Group	24

2.2.5	Influence of Branching Index BI	25
2.3	Fuels and Additives for Combustion Research	27
2.3.1	PRF/TPRF.....	28
2.3.2	Oxygenated Functional Groups	29
2.4	AI in Combustion Research	31
2.4.1	Applications for ML	31
2.4.2	Steps for ML Implementation.....	33
2.4.3	Historical Development of ML in Fuel Research	35
3	CHAPTER 3: SP MODELLING.....	37
3.1	Methodology	37
3.1.1	Selection of Oxygenated Compounds.....	38
3.1.2	Virtual Smoke Point VSP	40
3.2	Experimental Setup	42
3.3	Computational Setup	44
3.3.1	Dataset.....	45
3.3.2	Model Development.....	48
4	CHAPTER 4: RESULTS	49
4.1	Experimental Results.....	49
4.1.1	Impact of Alcohols.....	49
4.1.2	Impact of Ethers.....	52
4.1.3	Impact of Ketones	55
4.1.4	Threshold Sooting Index (TSI)	56
4.2	Computational Results:	59
4.2.1	ANN model	59
4.3	ML Models.....	63
4.3.1	ETR Model.....	63
4.3.2	XGBR Model	64
4.3.3	LightGBM Model	64
4.3.4	Nu-SVR Model	65
4.3.5	Support Vectors Machine	66
4.3.6	Random Forest	67

4.4	SHAP Analysis	68
4.5	ML Models Performance	69
4.6	ML Models Results:	72
4.7	Feature Importance and SHAP analysis for ML models	81
4.8	Model Evaluation: MLs vs. Deep Learning.....	84
5	CHAPTER 5: CONCLUSION.....	86
5.1	Conclusion	86
5.2	Future Work.....	88
6	REFERNCES	90
7	APPENDIX.....	104
	VITAE.....	109

LIST OF TABLES

Table 3-1: Physicochemical Properties of Selected Chemicals for SP Measurement....	39
Table 3-2: Initial investigation of the fuel blend	41
Table 3-3: Statistically analyzing dataset features for SP prediction based on wt%.....	46
Table 4-1: Significant characteristics of training parameters for ANN architecture.....	59
Table 4-2: Model performance metrics on TPRF80-20 blends	60
Table 4-3: Configuration of hyperparameter tuning for ML models and best-fit values	70
Table 4-4: Comparison of metrics performance for ML and deep learning models for predicting SP	84

LIST OF FIGURES

Figure 2.1: Stages of Soot Formation.....	11
Figure 2.2: Apparatus for SP Measurement, ASTM-D1322.....	14
Figure 2.3: Schematic of a standard ML development model.....	34
Figure 3.1: Flowchart for developing SP prediction model	38
Figure 3.2: Apparatus setup for SP Measurement, ASTM D1322	42
Figure 3.3: SP measurement steps	43
Figure 3.4: Steps for reading SP height for TPRF80-20 blended with cyclopentanone6%	44
Figure 3.5: Heatmap of Pearson Correlation Coefficients to represent SP relationship with input features	47
Figure 4.1: Impact of blending alcohols with TPRF 80-20 on SP	50
Figure 4.2: Impact of blending ethers with TPRF80-20 on SP	53
Figure 4.3: Impact of blending ketones with TPRF80-20 on SP	55
Figure 4.4: Impacts of adding alcohols(a), ethers(b), and ketones(c) with various percentages on TSI	57
Figure 4.5: Comparison of experimental and predicted SP values for training and testing data using Ann model.....	61
Figure 4.6: Plotting actual SP vs predicted SP for all and validation data using ANN model.	62
Figure 4.7: Predicted vs Actual values for training, testing, and entire dataset for ETR model	72
Figure 4.8: Predicted vs Actual values for training, testing, and entire dataset for XGBoost model	73
Figure 4.9: Predicted vs Actual values for training, testing, and entire dataset for LightGBM model	75
Figure 4.10: Predicted vs Actual values for training, testing, and the entire dataset for Nu-SVR model	76
Figure 4.11: Predicted vs Actual values for training, testing, and entire dataset for SVM model	77
Figure 4.12: Predicted vs Actual Values for Training, Testing, and Entire Dataset for RF model	79
Figure 4.13: Taylor Diagram for Performance Comparison of All Models.....	80
Figure 4.14: Permutation Method Illustration of Feature Importance on the SVM Model.....	82
Figure 4.15: SHAP Summary Plot for Feature Impact on SP based on SVM Model....	83
Figure 5.1: Soot tendency reduction for 14% oxygenated additives.....	87

LIST OF ABBREVIATIONS

ANN	: Artificial Neural Networks
OME	: Oxymethylene Dimethyl Ethers
SP	: Smoke Point
DCN	: Derived Cetane Number
NO_x	: Nitrogen Oxides
CI	: Compression Ignition
SI	: Spark Ignition
QSPR	: Quantitative Structure Property Relationship
MLR	: Multiple Linear Regression
MNR	: Multivariate Nonlinear Regression
FESI	: Fuel Equivalent Sooting Index
BI	: Branching Index
OESI	: Oxygen Extended Sooting Index
TSI	: Threshold Sooting Index
ASTM	: American Society for Testing Material
Nu-SVR	: Nu- Support Vector Regression

RF : Random Forest

ETR : Extra Tree Regression

SVM : Support Vector Machine

XGBR : Extreme Gradient Boosting Regression

LightGBR : Light Gradient Boosting Regression

MW : Molecular Weight

ABSTRACT

Full Name : Ruaa Ala Eldin Ageeb Abakar
Thesis Title : Implementation Of AI In Smoke Point Prediction
Major Field : Chemical Engineering
Date of Degree : December, 2024

The combustion of conventional hydrocarbon-based fuels exerts a profound adverse impact on the environment, primarily due to the emission of particulate matter such as soot. To mitigate these challenges, this study investigates advanced strategies for enhancing fuel combustion characteristics and curbing pollutant emissions. In particular, the research examines the effects of incorporating oxygenated functional groups—including alcohols, ethers, and ketones—into toluene primary reference fuel (TPRF) blends. Experimental results demonstrate that blending fuels with oxygenates significantly increases the smoke point (SP), which in return suppresses soot formation and promotes cleaner, more efficient combustion processes. A comprehensive dataset comprising 89 new SP measurements for TPRF with oxygenate highlights the pivotal role of oxygenates in reducing soot formation and enhancing fuel efficiency for advancing environmental sustainability.

This work implements advanced artificial intelligence (AI) approaches to forecast smoke point (SP) values of oxygenated fuel blends in order to develop fuel design methodologies. With an R-squared value of 0.98 and a Root Mean Squared Error (RMSE) of 0.84 and Mean Absolute Error (MAE) of 0.6 on the test set, Support Vector Machine (SVM) outperformed the other six machine learning (ML) models techniques in the training and testing of the framework across accuracy and generalization.

In addition, an artificial neural network (ANN) model demonstrated exceptional prediction accuracy on the test set, with an R-squared value of 0.97, RMSE of 0.23, and Mean Absolute Error (MAE) of 0.18. These trustworthy models eliminate the need for time-consuming experimental methods while enabling accurate evaluations of SP and the possibility for soot reduction that goes along with it. This approach opens the door for cleaner combustion techniques and environmentally sustainable fuels innovations by combining empirical experimentation with AI -driven computational modelling to give scalable and affordable ways for forecasting important fuel characteristics.

ملخص الرسالة

الاسم الكامل: رؤى علاء الدين عجيب ابكر

عنوان الرسالة: تطبيق الذكاء الاصطناعي في التنبؤ بنقطة الدخان.

التخصص: هندسة كيميائية

تاريخ الدرجة العلمية: ديسمبر 2024

يُشكّل احتراق الوقود الهيدروكربوني التقليدي تأثيرًا سلبيًا عميقًا على البيئة، وذلك بشكل أساسي بسبب انبعاث الجسيمات الدقيقة مثل السُخام. ولمواجهة هذه التحديات، يستكشف هذا البحث استراتيجيات متقدمة لتحسين خصائص احتراق الوقود وتقليل انبعاثات الملوثات. وعلى وجه الخصوص، يفحص البحث تأثير إضافة المجموعات الوظيفية المؤكسجة، مثل: الكحولات والإثيرات والكيونات، إلى خلطات الوقود المرجعي الأساسي من التولوين (TPRF) أظهرت النتائج التجريبية أن مزج الوقود مع المؤكسدات يؤدي بشكل كبير إلى رفع نقطة الدخان (SP)، مما يقلل من تكوّن السُخام ويعزز عمليات الاحتراق لوقود أنظف وأكثر كفاءة. كما يوفر هذا البحث مجموعة بيانات شاملة تتضمن 82 قياسًا جديدًا لنقطة الدخان لخلطات TPRF مع المؤكسدات، مما يبرز الدور الحيوي لهذه المؤكسدات في تحسين كفاءة الوقود وتعزيز الاستدامة البيئية.

تستخدم الدراسة الحالية تقنيات متقدمة للذكاء الاصطناعي (AI) للتنبؤ بقيم نقطة الدخان (SP) لخليط الوقود المحتوي على الأكسجين بهدف تطوير منهجيات تصميم الوقود. حقق نموذج آلة الدعم المتجه (SVM) أداءً متفوقًا مقارنة بالتقنيات التقليدية الأخرى في تطوير واختبار إطار عمل التعلم الآلي عبر عدة نماذج، حيث بلغت قيمة معامل التحديد (R^2) 0.98 وقيمة خطأ الجذر التربيعي المتوسط (RMSE) 0.84، ومتوسط الخطأ المطلق هو (MAE) 0.6 على مجموعة الاختبار. بالإضافة إلى ذلك، أظهر نموذج الشبكة العصبية الاصطناعية (ANN) دقة استثنائية في التنبؤ على مجموعة الاختبار، حيث بلغت قيمة معامل التحديد (R^2) 0.97، وقيمة خطأ الجذر التربيعي المتوسط (RMSE) 0.23، ومتوسط الخطأ المطلق هو (MAE) 0.18. تتيح هذه النماذج الموثوقة إجراء تقييمات دقيقة لنقطة الدخان وإمكانية تقليل السخام المصاحب لها، مع الاستغناء عن الحاجة إلى الطرق التجريبية التي تستهلك وقتًا طويلاً. يفتح هذا النهج المجال أمام تقنيات احتراق أنظف وابتكارات وقود صديقة للبيئة من خلال الجمع بين التجارب التجريبية والنمذجة الحسابية المعتمدة على الذكاء الاصطناعي لتوفير طرق قابلة للتوسع وذات تكلفة معقولة للتنبؤ بخصائص الوقود المهمة.

CHAPTER 1

INTRODUCTION

1.1 Background

Fuels are classified into two main categories: primary fuels, found naturally, such as coal, petroleum, and natural gas, and secondary fuels, derived from primary fuels through chemical reactions, and physical, or thermal processes, such as coke, gasoline, and coal gas. The applications of these fuels are vast, particularly in combustion equipment like furnaces, boilers, and combustion engines, which depend on solid, liquid, or gaseous fuels. Selecting the optimum fuel requires consideration of several factors, including quantity, quality, cost, environmental effects, storage, and heating value. Importantly, the properties of the fuel determine its quality and efficiency, which in turn influences its suitability for specific applications [1].

For decades, energy suppliers have relied heavily on primary energy sources such as fossil fuels for transportation, especially in hard-to-decarbonize industries. Despite this dependency, there is fast growing need for sustainable energies and renewable fuels due to the urgency of decarbonization. The use of synthetic fuels, such as oxymethylene dimethyl ethers (OMEs), shows great promise in reducing smoke formation and supporting low-carbon transport systems. Integration of such oxygenated fuels into existing infrastructures provides an essential comprehensive understanding of their thermochemical properties.

The challenges associated are due to the complex and variable composition of fossil fuels based on their origin and production processes [2]. Furthermore, biofuels that consist of oxygenated hydrocarbons, are emerging as a viable replacement for fossil fuels due to their low soot emissions and carbon neutrality. Despite their potential, many biofuel candidates remain under research regarding their thermochemical, physical, and combustion properties for transportation sectors.

Functional groups that contain large numbers of carbon, hydrogen, and oxygen require accurate characterization to understand their influence on the properties. Experimental analysis of all potential fuel candidates is impractical due to the complexity of the experiment and the cost associated with it. Traditional modeling approaches, based on quantum mechanical principles, are also resource-intensive and time-consuming [3].

Fuel properties, such as SP, derived cetane number (DCN), pour point, viscosity, flash point, calorific value, density, and octane number, have a significant effect on combustion performance. For example, a complex mixture of diesel fuel contains hydrocarbons with chain lengths of 9 to 25 carbon atoms, and surrogate fuels have been employed to simplify the complexity of these mixtures. However, predicting the behavior of fuels under extreme conditions, such as high pressure and temperature, remains as a significant challenge [2].

Emissions from burning fuels, such as nitrogen oxides (NO_x), Carbon dioxide (CO_2) and soot that consist of a carbonaceous particulate matter (PM), are critical concerns for both spark ignition (SI) and compression ignition (CI) engines.

SI engines, offering improved efficiency through strategies like stratified charge combustion, face challenging in managing PM and NO_x emissions due to the incompatibility of certain catalysts, such as the three-way catalyst (TWC), with stoichiometric conditions. Advanced approaches, such as low-temperature combustion (LTC) and advanced compression ignition (ACI), aim to suppress NO_x and PM emissions and improve thermal efficiency [4,5]. Fuels' properties, specially SP, provide an accurate indicator of fuels' sooting tendencies and can guide the engine design that can support emissions reduction and improve efficiency. For example, a study investigated the use of catalyzed diesel particulate filters (CDPF) and diesel oxidation catalysts (DOC) demonstrated over 90% reduction in PM and particle number emissions [6]. Understanding SP property is essential for environmental sustainability, reducing air pollutants like CO₂ contributes to mitigating climate change and its associated impacts, such as global warming and fire growth [7].

Understanding of the physical and chemical properties of fuels is necessary for numerical simulations of combustion phenomena, designing fuel, and optimizing engine in internal combustion systems and gas turbines. The accuracy of these simulations depends on precise property estimation, making the development of efficient predictive tools are critical [8]. The implementation of advanced technology, including AI and ML, is a transforming industry by improving efficiency, reducing costs, and minimizing human involvement. The industrial evolution era over the history has progressively moved from manual processes to programmable control systems and now to intelligent systems, integrating technologies such as edge computing, internet of things, and ML [9, 10].

ML has emerged as a powerful tool for analyzing complex datasets and making predictions possible with minimal human supervision. By leveraging large datasets, ML algorithms can accurately estimate fuel properties, even for complex mixtures under varying conditions of pressure and temperature [2]. As a subset of AI, ML enables computers to learn patterns from data and make decisions without being explicitly programmed. Arthur Samuel first defined ML in 1959 as the ability of machines to learn without direct programming [11]. Various deep learning techniques, including ANN, supervised learning methods, and unsupervised learning methods, have been explored for fuel property prediction. Among most of AI models, ANN are particularly effective due to the ability of such deep learning model to handle nonlinear and complex relationships between input features and output which results in reducing data complexity through hidden layers and aggregation processes [3,12].

ANNs have demonstrated superior accuracy, efficiency, and cost-effectiveness compared to traditional models, making them a valuable tool for predicting properties like SP [13]. Considering the implementation of ML into the field of fuel research facilitates the discovery of new biofuel/ oxygenated candidates and accelerates the development of sustainable fuels tailored to meet future energy demands. This work focuses on leveraging six ML models and an ANN to predict fuel combustion property efficiently, enabling advancements in fuel design and engine optimization.

1.2 Problem Statement

Prediction of oxygenated fuel properties is not accurate as the complexity of the functional groups increases due to the increase in the nonlinearity of the fuel properties.

Therefore, this limitation highlights the need for more accurate prediction methods that identifies the relationship between variety of the functional groups and SP that captures the non-linearity between the data which other methods struggle to capture it effectively.

An advanced prediction method such as implementing of ANN model, which is capable to handle this nonlinearity, is highly recommended for investigation.

Depending on the experimental data that the study relies on, primary reference fuel (PRF) and toluene primary reference fuel (TPRF) were used as base fuels, blended with a variety of oxygenated fuels, including five alcohols, two ketones, and four ethers, to validate the prediction of combustion properties using functional groups as descriptors. SP is a critical property for determining the fuel tendency to produce soot, and combustion efficiency serves as a main focus. Additionally, this research explores the capability of ML models in predicting highly nonlinear relationship between descriptors and SP, with an emphasis on the highly effective functional group in predicting SP. In the fuel design part, this research aims to predict the functional group percentage for a targeted SP.

With the power of optimization algorithms, the aim to enhance the prediction of blended oxygenated fuels is a major to refine the blending process and improve the overall fuels' performance. Such a pathway addresses the challenges of emission reduction and enhancing combustion efficiency.

1.3 Aims and Objectives

- To experimentally measure the SP of oxygenated fuels containing alcohols, ketones, and ethers using PRF/TPRFs as base fuels.
- To calculate the sooting potential of fuels, threshold sooting index (TSI) techniques have been measured for the blended fuels, in addition to investigating which of the additives has the highest soot reduction based on the SP and the molecular weight of the blend
- To design, implement and validate an ANN model for accurately predicting SP value for oxygenated fuel blends.
- To compare ANN model results with RF, ETR, SVM, Nu-SVR, XGBR and light-GBR and identify which of these models have the best results based on accuracy which define the model capability to predict SP and generalization which depends on the model's performance on the unseen data.
- To validate and refine the optimal blends for fuel design that align with environmental standards.

CHAPTER 2

LITERATURE REVIEW

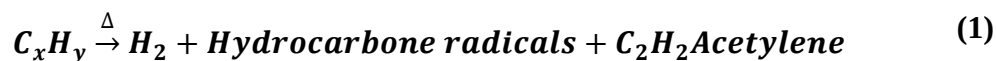
The combustion process is a highly exothermic, redox reaction between fuel and atmospheric oxygen. Such a reaction releases heat and produces visible flames in addition to the product. The dependency on fossil fuels spans a cross variety of sectors, from power generation, where the three types of fossil fuels coal, gas, and oil are consumed extensively to meet growing electricity demands, to heavy industries such as cement and steel manufacturing rely on these energy sources to drive high-temperature processes. However, the transportation sector emerges as the largest consumer, accounting for over 95% of fossil fuel usage, with renewable energy sources contributing only about 5%, underscoring their pivotal role in global energy systems [14,15]. This fact emphasizes the critical role of fossil fuels in powering vehicles, ships, and aircraft. Central to their utilization is the combustion process—a highly exothermic redox reaction between the fuel and atmospheric oxygen. Such reaction generates substantial heat, produces visible flames, and releases various byproducts[16]. Such a dependency comes with significant environmental and health consequences. The emissions out of the combustion process, lead to greenhouse gas emissions, including carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), which drive global warming. It also produces pollutants like sulfur oxides (SO_x), nitrogen oxides (NO_x), and particulate matter, all of which degrade air quality. Among these, soot, primarily composed of black carbon, is this study's concern.

Soot is considered the 2nd contributor to global warming. With the high thermal absorbent capability, soot increases thermal radiation that causes flame spread and fire growth.

2.1 Soot Formation

Carbon-based particles known as soot are created when hydrocarbon fuels are not completely burned or undergo pyrolysis. The negative impacts of soot emissions on agriculture, population health, the condition of the atmosphere, and both the worldwide and local climate are well documented. Soot formation is a complex chemical, and physical process that involves multiple stage [17,18]:

1. **Thermal Decomposition:** Under high temperatures, the hydrocarbon fuels decompose, leading to the production of hydrogen, acetylene and smaller hydrocarbon radicals. The main components of soot precursors are provided by **equation (1)**. This is the breakdown:



2. **Aromatic Growth (Precursor Formation):** A key stage in the production of soot is the synthesis of aromatic structures, which serve as the building blocks for the production of polycyclic aromatic hydrocarbons (PAHs) and, eventually, soot particles. This phase includes:

- a) **Benzene Ring Formation:** Acetylene (C_2H_2) or propargyl radicals (C_3H_3) combine with tiny hydrocarbon radicals to generate benzene (C_6H_6), the first aromatic structure. The elevated temperature and the presence of reactive intermediates in the flame promote this reaction as shown by **equation (2)**.



- b) **Formation of Naphthalene ($C_{10}H_8$):** The benzene ring develops into naphthalene, a two-ring aromatic hydrocarbon, through a series of interactions involving acetylene and other precursors via **equation (3)**. This is the first stage of PAH development.



- c) **Additional PAH Growth:** By adding more acetylene and recombining PAH fragments, naphthalene and related structures continue to grow, forming larger PAHs like pyrene ($C_{16}H_{10}$) and phenanthrene ($C_{14}H_{10}$) as in **equation (4)**, and subsequently more complex molecular weight MW type that nucleate to produce soot particles.



3. **Nucleation:** PAHs combine to form initial soot particles as in **equation (5)**, which signifies the change from gaseous to solid-phase particulate matter.



4. **The surface Growth:** Soot fragments develop by adsorbing gaseous hydrocarbons and adding carbon atoms, via **equation (6)**, which is made possible by processes such as the hydrogen-abstraction-carbon-addition (HACA) process.



5. **The aggregation and Coagulation:** Individual soot particles collide and stick to one another to form larger fractal-like aggregates provided by **equation (7)**; this stage influences the shape and size distribution of soot particles.



6. **The oxidation process:** In oxygen-rich areas of the flame or combustion circumstances, soot particles undergo oxidation as in **equation (8)**, which reduces their mass and changes their nanostructure, which can have a substantial effect on the emission characteristics of soot.



The summary of the key stages of soot formation is presented in **Figure 2.1**

Soot Formation Process

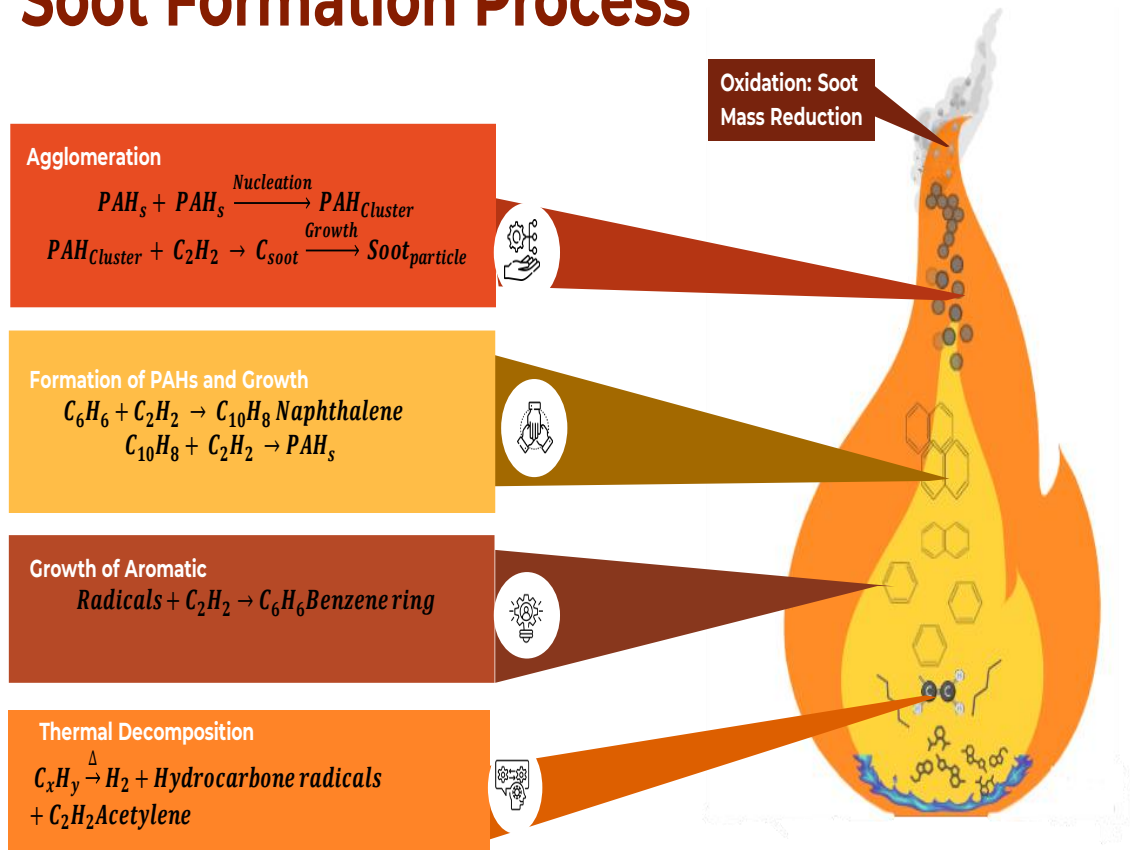


Figure 2.1: Stages of Soot Formation

2.2 Combustion Properties: Smoke Point (SP)

The SP serves as a pivotal metric for evaluating the fuel propensity to soot, including jet fuel, diesel, and kerosene. It represents the critical temperature-related threshold at which particulate matter (PM), or soot begins to form because of incomplete combustion [19]. SP and its relation to soot will be investigated in the below sections.

2.2.1 SP as a measure of Sooting Index

The SP of a fuel is conventionally defined by the maximum laminar flame height—often described as a "candle-like" flame at which no visible smoke is emitted from the flame tip. Soot formation generally occurs in the lower region of the flame, where fuel decomposition and intermediate species are present. The soot then moves upward by ascending air and oxidised in the upper part of the flame. The balance between soot generation and oxidation in laminar flames operating at their smoke-point conditions guarantees that soot is fully oxidised before exiting. In contrast, soot oxidation dominates and produces very little soot emissions for flames burning at or below their SP height [20,21]. Several factors affect the production of soot, but residence time is a crucial one. In the lower flame area, where undecomposed fuel and intermediates are still accessible, the percentage of fuel converted to soot rises as fuel residence duration increases. Higher radiative heat losses per unit volume of soot and higher soot volume fractions are the results of longer residence durations. Both the residence time and radiative heat loss scale with the fuel supply rate (or flame volume) [22].

As the fuel supply rate increases, radiative heat loss intensifies, leading to increased flame cooling. Eventually, excessive cooling causes the flame tip to quench before complete oxidation of soot and other products of incomplete combustion can occur [20]. Lower laminar flame heights at the SP, shorter residence periods, and more noticeable radiative heat losses, all of which cause quick cooling and early flame extinction are characteristics of fuels with increased sooting tendencies [22]. At elevated combustion temperatures, the nature of flame diffusion, whether turbulent or laminar, plays a crucial role in determining the reaction mechanism [23]. In laminar diffusion, an increase in the fuel flow rate leads to the flame reaching its SP, marked by a significant plume of soot breakthrough above the flame [24]. The SP, a temperature-related parameter, indicates the threshold at which fuel generates PM or soot due to incomplete combustion. This parameter serves as a characteristic feature for jet fuel, diesel, and kerosene, varying across different fuel types with a range of 300°F to 550°F [25].

Utilizing the ASTM-D1322 Standard [26], the SP lamp apparatus is employed to measure the SP value **Figure 2.2**, by observing the highest flame in mm without observing the soot mist. The average height is then used as input data for the model. All tools were maintained clean via using n-heptane solution and ionized water. Eventually, an estimated value of each fuel mixture's propensity to soot was obtained by measuring the SP as they inversely proportion [27].

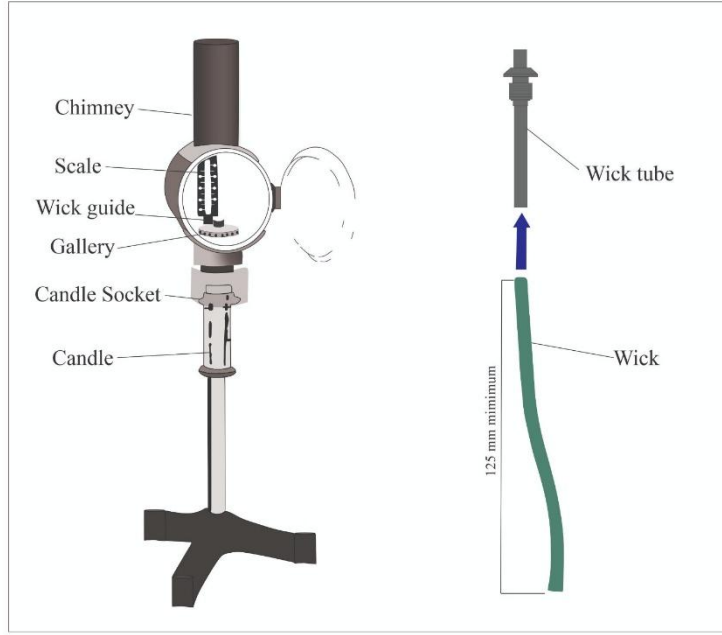


Figure 2.2: Apparatus for SP Measurement, ASTM-D1322.

The relationship between soot and SP is inversely proportional with a higher SP flow rate corresponding to a lower tendency for sooting [28]. To assess the sooting potential of fuels, various techniques have been developed. The TSI, proposed by Calcote and Manos (1983) [29], has extensively applied while gaining significant recognition. TSI is the numerical value that indicates a fuel's propensity to soot, with higher values reflecting a greater tendency to produce soot. TSI is calculated based on the MW and the SP using **equation (9)** as follows:

$$TSI = a \left(\frac{MW}{SP} \right) + b \quad (9)$$

From the above equation 9 the notations ‘a’ and ‘b’ are constants specified to the SP apparatus and MW is molecular weight of the fuel. Such an equation highlights the importance of accurately predicting SP to assess the sooting tendencies of various fuels, ultimately leading to the development of cleaner combustion technologies.

Based on Olson et al. work, the TSI was measured for 103 hydrocarbons, revealing the important aspects of soot formation mechanisms [27]. Further research done by Gill and Olson showed that the TSI for fuel blends can be accurately predicted using an additive rule as in **equation (10)**, highlighting the relationship between the mixture components and their sooting tendencies [30].

$$TSI_m = \sum_c x_c * TSI_c \quad (10)$$

From **equation (10)**, m represents the mixture and x_c represent the mole fraction of component c. Furthermore, variety of techniques developed to addresses the limitations of traditional sooting indices, these techniques typically do not consider the effects of the presence of oxygen atoms in fuel molecules [31], oxygen extended sooting index (OESI) [32] is used as a useful method to study the effect of oxygenated groups such as alcohols, ketones and ethers in both aromatic and non-aromatic surrogate fuels. OESI is calculated using **equation (11)** below:

$$OESI = a^* \left(\frac{C + \frac{h}{4} - \frac{o}{2}}{SP} \right) + b^* \quad (11)$$

The parameters used in this equation stands for: C the carbon, representing the number of carbon atoms in the fuel molecule that play a direct role in soot formation during combustion. H stands for hydrogen, indicating the H atoms that affect the combustion process. O stands for Oxygen, referring to the O atoms in the fuel molecule that enhance combustion and reduce soot formation by promoting oxidation processes that break down soot precursors. and a^* and b^* are constant for the device. The value of SP is integral to both TSI and OESI which have similar values for non-oxygenated surrogates.

It is important to note that the SP is relatively user-dependent, which leads to significant uncertainties in its measurement. Additionally, the SP of aromatic compounds is notably low and challenging to measure, potentially limiting the utility of this parameter in analyzing practical fuels [33]. Research, improved sooting indices through developing technique such as the yield sooting index (YSI) [33–37], YSI estimates sooting tendencies of the fuel through measuring the maximum volume fraction V_{fmax} of the soot through laser-induced incandescence (LII). Using constants $\acute{a}$ and $\acute{b}$, the V_{fmax} value is converted into an apparatus-independent index, as described in **equation (12)** as follow:

$$YSI = \acute{a} * V_{fmax} + \acute{b} \quad (12)$$

Another method for measuring fuel propensity to soot is the micro pyrolysis index (MPI). This method was developed with several key features: (i) it is independent of oxygen levels, addressing the significant variations in sooting tendencies that resulted from changing the oxygen concentration.

(ii) it requires only a small volume from the sample, making it suitable for laboratory-scale applications. (iii) it improves repeatability compared to traditional methods; and (iv) it uses standard laboratory equipment, facilitating its adoption across multiple research settings. MPI measure requires a 20 μL of fuel, this fuel is then pyrolyzed over a bed of $\text{R-Al}_2\text{O}_3$ beads at 850 $^\circ\text{C}$. The carbon deposited on the beads is then oxidized in a clean reactor using temperature-programmed oxidation (TPO) to quantify the soot formed. This method provides a reliable measure of a fuel's sooting tendency, independent of parameters like flame temperature or stoichiometric oxygen/fuel ratios, which can significantly influence results in other sooting tendency assessment techniques [38].

While fuel equivalent sooting index (FESI) incorporates the sooting tendencies measured through YSI, which is based on the soot volume fraction produced under specific combustion conditions, and OESI, which adjusts for the impact of oxygenated compounds on soot reduction. By integrating these concepts, FESI provides a comprehensive measure of a fuel's propensity to form soot, accounting for both molecular structure and oxygen content, making it a more versatile and accurate indicator of sooting tendencies for complex fuel blends, especially those containing oxygenates [39,40]. FESI value expressed based on **equation (13)** as follow:

$$FESI_y = \frac{A * V_{fmax} - \sum_c x_c * FESI_c}{x_y} \quad (13)$$

The samples of **equation (13)** defined as follow:

‘A’ represents an apparatus-dependent constant, while x_c and $FESI_c$ denote the mole fraction and Fuel Equivalent Sooting Index (FESI) of the c^{th} component in the blend, respectively. The x_y refers to the mole fraction of the fuel component whose sooting tendency is being evaluated, denoted as $FESI_y$. The hydrogen content of pure compounds, determined from their molecular formulas, is linked to soot production during combustion, with higher hydrogen content yielding less soot. Furthermore, the aromatic content of fuel significantly influences soot formation, with higher aromatic content generally correlating with increased soot formation [41].

The emission of particulate matter (PM) and soot affects human inhalation system, engines, and the environment. Due to that pollution, Environmental Protection Agency (EPA) has its standard for acceptable percentage of soot emission [5]. Many studies were involved in improving fuel properties through additives to decrease the amount of PM and NO_x . Data were collected from pure hydrocarbon and oxygenated fuel to form models for predicting SP which used in calculating the TSI [5,42].

An input of (jet fuel, kerosene and gasoline and their surrogate) with those additives containing alcohol and ethers, MW along with branching index (BI) of organic groups of olefinic, naphthenic and paraffinic of methane, ethane, propane, ethene, ethanolic OH/O and aromatic were included. The incompetence of multiple linear regression method to capture the data for SP model, ANN model with high accuracy is used [42]. Previously many approaches have been employed to assess the sooting tendencies of oxygenates.

The first approach involved creating flames from pure oxygenates and measuring the smoke height, defined as the maximum flame height of the test compound that did not emit soot from its tip. This technique had been utilized for seven long-chain alcohols and carboxylic acids, as well as for five small alcohols [43,44].

The second approach added oxygenates to a heavily sooting base flame and measured the consequent reductions in smoke height. This method was advantageous as it was more closely related to the goal of minimizing soot formation in engines. It had been applied to 24 diverse oxygenates [45], using either a heptane/toluene mixture or a commercial diesel fuel as the base fuel. A third approach involved introducing small amounts of oxygenates to a lightly sooting base flame and measuring the subsequent increases in soot concentration. The primary benefit of this method was that the additive concentrations ~1000 ppm were too low to impact soot concentrations through indirect mechanisms such as base fuel dilution or flame temperature changes. Therefore, the observed sooting tendencies were predominantly influenced by the direct chemical effects of the oxygenate structure. Additionally, the method's experimental convenience allowed for the characterization of hundreds of test compounds within a reasonable timeframe [36].

The SP range for n-heptane and isooctane is beyond ASTM-1322 values therefore PRF is also beyond the range. Hence Haas.et.al [46] utilized the TSI linear by mole blending approach to predict the SP for known MW fuels with a minimal sooting propensity using extrapolation. In this work we aim to validate the model using the VSP data collected experimentally.

2.2.2 Methods to Improve SP

Enhancing SP reduces the production of soot, which minimizes its detrimental effects on atmosphere quality and consequences to the environment and global warming, in addition to the effects on human health. It also improves combustion efficiency. Main strategies to deal with these issues are examined in this section.

Fuel pre-treatment or catalytic processes: Researchers as Eliaz [47] study the use of catalytic systems to chemically pre-treat fuels, with an emphasis on oxidizing polycyclic aromatic hydrocarbons (PAHs) and disrupting the intricate structure of hydrocarbons. In high-pressure combustion systems, this catalytic initial treatment uses elements like platinum to promote reactions like paraffin dehydrogenation, which forms olefins and reduces soot precursors. These methods successfully lower soot generation and raise the fuel's SP by changing the fuel's chemical structure before combustion.

Optimizing combustion conditions is another technique used to improve the SP. Soot production can be considerably decreased by fine-tuning the operational parameters of combustion systems, such as ensuring the air-fuel mixture is proper, managing the temperature during combustion, and modifying the pressure levels.

To minimize the production of unburned hydrocarbons, optimum air-fuel mixing guarantees an appropriate proportion for full combustion. The problems of high temperatures, which encourage the generation of nitrogen oxides (NO_x), and low temperatures, which lead to incomplete combustion, are avoided by maintaining the temperature of combustion within an ideal range.

Furthermore, sustaining high pressure increases air mixing and fuel atomization, which boosts combustion efficiency. By raising the SP and minimizing the incomplete combustion zone, these modifications successfully lower soot emissions [48,49].

Furthermore, adjusting the **fuel composition method** is another step that can be implemented to improve fuels' SP, by changing the composition structurally to greatly decrease the fuel's propensity for making soot. One of the main processes is lowering the aromatic content, which reduces the production of polycyclic aromatic hydrocarbons (PAHs), a contributor to soot formation. Combustion characteristics are also enhanced by reducing the carbon-to-hydrogen ratio and raising the paraffinic concentration. These fuels have a higher SP and significantly lower soot emissions because they produce fewer carbon residues when they burn [50–52].

Blending Oxygenated Fuels: As a final strategy to lower soot emissions, this study investigates the blending of oxygenated chemicals into fuels. This method reduces the incomplete combustion zones where soot precursors usually form by promoting more complete combustion through an increase in the oxygen-to-carbon (O/C) ratio of the fuel. It is an efficient method of lowering emissions because the increased oxygen availability interferes with the chemical processes that cause soot particles to develop. This technique shows promise as a workable alternative for cleaner fuel applications since it not only improves SP and combustion efficiency but also dramatically reduces soot production. [53–55].

2.2.3 Influence of Functional Groups on Combustion Properties

Crude oil represents one of the most chemically intricate substances found in nature, several hundred thousand distinct molecular entities varying in size, structure, and functionality exist within a single drop of oil. In such a complex mixture, a functional group is defined as a specific arrangement of atoms within a molecule that dictates its physical properties, chemical behavior, and reactivity. [56,57].

Convenient fuels like gasoline and diesel are made up of a large number of hydrocarbon molecules, usually, gasoline consists of 6th to 7th groups [58]. These hydrocarbons often have a restricted range of functional groups, despite their complexity. Such functional groups are essential in controlling the fuel's combustion kinetics, which in turn determines important performance attributes like emissions characteristics, ignition quality, and flame stability [59]. The amount and nature of functional groups that are present in a compound have a significant impact on its thermochemical properties (heat capacity [60], heat of formation [61], and flash point [62–65]), physical properties such as density and viscosity [66], and different combustion characteristics (cetane number, flame speed, SP and sooting tendency [25,28,67–69]).

There is a strong relationship between a fuel's physical and chemical behavior and the distribution of functional groups within it [70,71]. By providing crucial insights into fuel efficiency optimized performance, Dooley's work[72] further emphasizes the crucial significance that functional group composition plays in shaping these qualities.

It has been observed that the component molecules are essential for controlling chemical reactions, independent of the parent surrogate molecule. The most common structural units in paraffinic hydrocarbon fuels are the CH_3 and CH_2 groups.

According to Wn et al. [70], the CH_2/CH_3 ratio and the metric $\frac{\text{CH}_2}{\text{CH}_3} * (\text{CH}_2 + \text{CH}_3)$ are essential for determining the low-temperature alkyl peroxy radicals' reactivity in paraffinic fuels. Additionally, the CH_2/CH_3 ratio in n-paraffin fuels, is a significant factor in determining the ignition delay especially within the boiling range of diesel fuels [71,73].

Many studies have investigated using the functional group's additive technique to estimate reaction rates based on alkyl and free radical groups [74]. To estimate reaction rates for a range of reaction sets used in kinetic simulations, Sumathi et al. [75–77] expanded Benson's group additivity. In a similar vein, Pepiot-Desjardin et al. [78] applied group additivity techniques to forecast the sooting patterns among several fuel types. The TSI in gas engine combustors was estimated by Yang et al. [79] using fundamental compositional data from jet fuels. Mehl et al. [80] suggested defining fuel surrogates for extremely heavy molecular substances by employing pseudo-components that mimic the reactive groups present in the desired fuel. Furthermore, structural moiety contribution frameworks have been developed to analyze the sooting tendencies of gasoline and diesel [78] and to estimate the smoke points of pure hydrocarbons [81].

The fact that various functional groups produce distinct sooting behaviors is further supported by experimental data [31].

These results demonstrate that the fundamental basis for predicting crucial combustion phenomena is the functional groups present in fuels. Since a fuel's behavior is largely determined by its chemical structure, studying functional groups is essential to improving our knowledge of combustion kinetics and emissions.

2.2.4 Functional Group

Naphthene's, olefins, iso-paraffins, n-paraffins, and aromatics are the five molecular groups that make up the majority of motorized fuels, including petrol, diesel, kerosene and jet fuel. Furthermore, oxygenates are frequently added to fuels in order to improve combustion and lower emissions. These reactive groups can be determined explicitly from the chemical structure and composition information for both pure hydrocarbons and hydrocarbon-oxygenated mixtures. As an illustration for ketones, cyclopentanone (MW= 84 g/mol) consist of 5 paraffinic CH₂ (MW= 14 g/mol), 1 carbonyl group (MW = 28 g/mol). The percentage weight of these group is 67% and 33% respectively. For a mixture of TPRF 80-20 (MW = 105.72 g/mol) consist of (toluene 20% vol, iso-octane 64% and n-heptane 16% vol) the weight percentage can be identified with a knowledge of density (0.726 g/cm³) to determine the functional group content as follow: paraffinic: CH₃ 48.57%, CH₂ 18.05, CH 6.96% and C 6.43% and Aromatic 19.99%. Weight percentage (wt.%), as compared to mole percentage (mol.%), was used to quantify the functional groups. This decision was taken because, in some situations, mol.% numbers may give a somewhat inaccurate overview of the contributions made by functional groups to the overall composition. For instance, toluene has one aromatic C-CH (single ring) and a single CH₃ paraffinic group.

Each of these groups makes up 50% of the molecule in terms of mole percentage (mol.%). Nevertheless, even though the paraffinic CH₃ group has a much lower molecular weight than the aromatic ring, it is disproportionately highlighted in this computation. The computed mol.% values for the paraffinic CH₃ groups in n-heptane-toluene blends are found to be 15–20% higher than the equivalent weight percentage (wt.%) values, highlighting this disparity.

2.2.5 Influence of Branching Index BI

Under various circumstances, functional groups are essential for regulating the characteristics of combustion. Nonetheless, research indicates that isomers containing identical functional groups can differ greatly in low-temperature reactivity and octane or cetane values. For example, because of the different configurations of their functional groups, 2-methylheptane and 3-methylheptane have different combustion properties [82–86]. Abdul Jameel et al. developed the idea of the BI as an extra input parameter to describe the level of branching in a molecular structure [87].

An indicator known as the BI was developed to measure how functional groups are arranged within molecules. A Position Index (PI) is used to identify the position of critical functional groups, like alkyl substitutions, within the BI architecture. The degree of branching in the carbon chain positively influences octane number (ON), and negatively affects cetane number (CN) for pure substances and blends, as has been extensively published in research [88,89]. In particular, a greater level of branching (i.e., more alkyl substitutions) that associated with a drop in CN [87] and a rise in ON [90], when all other variables are held unchanged.

There is still not enough systematic study to measure the effect of branching on the combustion and thermochemical characteristics of fuels, despite the fact that many studies recognize its influence. Several research [68] have tried to use the CH_3/CH_2 ratio to depict branching in iso-paraffins, but this metric does not take into consideration branching in compounds like 2,4-dimethylbutane, iso-butane, and neo-pentane that lack CH_2 groups. Furthermore, the positional effects of alkyl substitutions on CN are not reflected in the CH_3/CH_2 ratio. For instance, 3-methylheptane and 2-methylheptane have the same CH_3/CH_2 ratio, but they have very distinct DCNs, which are 45 and 52.6, respectively.

Furthermore, because n-paraffins are unbranched, this ratio gives them a positive branching significance, which should ideally be zero. The degree of branching in a molecule or hydrocarbon mixture is generally indicated by the quantity of paraffinic CH_3 groups available; a higher number of CH_3 groups indicates a fuel that is more strongly branched. In contrast, the linearity of the structure is shown by the amount of paraffinic CH_2 groups. Due to the faster rate of low temperature peroxy radical isomerization processes between CH_2 groups, which encourage radical chain branching and ignite, highly linear fuels typically have more CN and lower ON [91,92]. The default setting of BI is taken to be 0 for olefinic, n-paraffinic, and straight chain molecules.

Equation (14) defines BI for iso-paraffins.

$$BI = \frac{C^b + PI_c}{C^l} \quad (14)$$

C^b represent the carbon atom branched from the longest chain, PI_c position index of C^b and C^l carbon atom of the longest chain. PI_c has a variety of values based on the position of C^b as follows:

$PI_c = 0, 0.5$ and $(0.5+0.5)$ based on the C^b connected to the position of the outermost, 1 position far from the outermost and 2 position far from outermost respectively. For detailed information 2,3-dimethylpentane BI's calculated in accordance with **Equation (14)** where PI_c value 0.5 for the 2nd C atom eventually the value of $BI = \frac{2+(0+0.5)}{5} = 0.5$.

Taking into account an input feature such as BI, allows one to differentiate 2,4-dimethylpentane from 2,3-dimethylpentane. However, compounds with naphthenic rings or aromatic are unable to have their BI calculated using **Equation (14)**, thus another calculation is required where ring effect is considered referring to [93].

2.3 Fuels and Additives for Combustion Research

The combustion process, despite the engine types, is significantly influenced by the Fuels and their properties. The physical properties significantly impact atomization quality, spray speed of evaporation and engine efficiency. Furthermore, the molecular composition of fuels greatly impacts environmental emissions, both in type and quantity. For example, fuels with high aromatic content tend to increase emissions, whereas adding oxygenated compounds reduces them. Ultimately, both fuel types and additives have a substantial effect on fuel and combustion characteristics[53,94–96]. This section will provide a detailed investigation of fuel bases and additives.

2.3.1 PRF/TPRF

Developing and validating chemical kinetic models for gasoline-like fuels is essential in fuel research. Given the complex composition of gasoline and other real fuels, creating a kinetic model for each individual component is highly challenging. To simplify this process, surrogate fuels are often formulated to mimic the key characteristics of the actual fuel. These characteristics typically include combustion properties, such as ignition delay and flame speed, as well as physical properties like MW, H/C ratio, viscosity, density, and the distillation curve. One of the most important findings is that no single surrogate fuel can perfectly replicate all these characteristics, eventually, a selection of the right surrogate requires careful consideration. Extensive discussions on surrogate fuel formulation strategies are found in the literature[97–100].

Historically, PRFs, defined by their research octane number (RON), were widely used as surrogates in combustion research.

PRFs, the blends of n-heptane and iso-octane, provide a simplified model for studying gasoline's auto-ignition behavior under a controlled condition. Extensive research has focused on developing chemical kinetic models for PRFs, making them essential for understanding fuel performance, especially in relation to engine knock resistance. In recent years, research has expanded to include complex fuel blends and additives, such as toluene, which is frequently used in TPRFs. TPRFs allow researchers to examine the effects of aromatic compounds on octane numbers and knock resistance. Although toluene has a high-octane rating of RON 120, its blending effects can vary, showing both synergistic and antagonistic behaviors based on the composition of the base fuel.

While toluene enhances octane levels, its effectiveness is significantly dependent on the chemical composition of the base fuel, especially its aromatic content. In summary, PRFs serve as a basic model for studying fuel behavior, whereas TPRFs provide more detailed insights into the role of aromatics like toluene, in influencing fuel performance [101–103].

2.3.2 Oxygenated Functional Groups

Blending fuels with oxygenated compounds, including alcohols, ethers, and ketones, has proven effectiveness in improving fuel properties, especially by increasing the smoke point. Oxygenated functional groups promote more complete combustion by enhancing the oxidation of hydrocarbons, thereby reducing the formation of unburned carbon particles (soot). As a result, fuels blended with oxygenated compounds tend to have higher smoke points and produce fewer particulate emissions.

Various studies have explored the effects of blending oxygenated compounds with surrogate fuels to optimize fuel quality and reduce soot emissions. For example, ethanol (EtOH) and dimethyl ether (DME) have been shown to significantly reduce sooting tendencies compared to pure hydrocarbons[104–113].

Other oxygenated compounds, such as n-pentanol, have also demonstrated improved combustion properties, including increased smoke points, when blended with conventional fuels like diesel[114,115]. Wei et al. found that blending n-pentanol with diesel not only increased the SP but also reduced hydrocarbon emissions, although it resulted in a slight increase in nitrogen oxides (NO_x) due to the presence of the alcohol (OH) group[116].

In addition to alcohol, ketones have recently gained attention for their ability to improve fuel efficiency and reduce emissions. Patil et al. examined the effects of blending various ketones, including cyclohexanone and 2-butanone, with gasoline and observed significant increases in the SP and reductions in soot formation[117]. The use of ketones in combustion engines, particularly those with common rail injection systems or direct injection engines, has been shown to reduce fuel consumption and enhance overall fuel performance[118].

The successful blending of these oxygenated compounds with conventional fuels demonstrates the potential for creating cleaner-burning fuels with higher smoke points. However, the complexity of fuel blends and the variety of oxygenated compounds available necessitate a comprehensive understanding of their interactions and impacts on combustion properties. Continuous research is essential to further optimize fuel formulations and reduce emissions[88].

2.4 AI in Combustion Research

In combustion research the use of AI has emerged as a powerful tool, enabling to accurately predict fuel properties [119,120]. Implementing AI approaches such as ML and deep learning in predicting engine performance and fuel design optimization is highly investigated [81,120–124]demonstrating the estimate experimental results which will lead to reduce physical testing with model can achieve high accuracy such as 97% and 98% respectively[25,125]. Further discussion will be included in this section.

2.4.1 Applications for ML

ML is a multidisciplinary field supported by numerous areas of study and has emerged as a powerful tool across numerous domains. The complexities associated with measuring the quantitative and qualitative aspects of thermochemical and physical fuel properties often require time-consuming and expensive testing. ML techniques, effectively address these challenges with their ability to learn from existing datasets and processes, by simplifying the prediction process and minimizing the dependency on extensive experiments [126]. The optimization of engine and fuel design requires a detailed understanding of the physical and chemical properties of fuels. This understanding is essential for conducting accurate numerical simulations of combustion phenomena in internal combustion engines and gas turbines. The accuracy of such simulations depends on reliable property estimation, highlighting the importance of developing efficient predictive tools [127]. Advancements in AI and ML are revolutionizing industries by enhancing efficiency, reducing costs, and minimizing human involvement.

Over time, industrial evolution has gradually shifted from manual operation to programmable control systems, and in recent time, to intelligent systems that integrate advanced technologies like edge computing, the Internet of Things (IoT), and ML [128,129]. These technologies enable industries to analyze complex datasets and make predictions with limited supervision. Various ML techniques, including ANNs, supervised learning, and unsupervised learning, have been applied to predict fuel properties. Among these methods, ANNs have demonstrated superior performance due to their ability to handle complex, nonlinear relationships and reduce data complexity through hidden layers and aggregation processes [62,130] [32]. ANNs have proven to be more accurate, efficient, and cost-effective than traditional models, making them particularly valuable for predicting properties like SP and DCN [131].

The integration of ML into fuel research has facilitated the discovery of biofuel candidates and accelerated the development of sustainable fuels tailored to meet future energy demands. The success of ML models relies on the quality of the database, which integrates experimental and computational findings, and the selection of descriptors that effectively represent the system.

Descriptors such as atom type electro-topological State Index (AET), BI, coulomb matrix (CM), functional group molecular descriptors (FGMD), molecular graph (MG), and MW provide critical input features. Advanced analytical tools like Fourier transform infrared (FTIR) spectroscopy, ASTM distillation curves, Nuclear Magnetic Resonance (NMR) spectroscopy, and dielectric spectroscopy further refine these descriptors, enabling accurate prediction of fuel properties [132,133].

Since hydrocarbon fuels comprise hundreds of distinct molecules of varying shapes and chemical types, their composition is often represented by grouping individual molecules into defined chemical moieties or functional groups. This approach simplifies the prediction of fuel properties, reducing experimental costs and complexities [134].

2.4.2 Steps for ML Implementation

For the purpose of predicting combustion quality (i.e., SP and DCN), various techniques and interactions have been established for n-paraffins[135], pure petroleum products [136–139], pure fuel blends, oxygen-rich hydrocarbons [140–143], fuels made from diesel [140,144–147], artificial types of diesel, biofuels, additives for biofuels and biodiesel [148,149]. A wide variety of study fields support the establishment of the multidisciplinary discipline of ML.

Because time-consuming and expensive tests are required, measuring the quantitative and qualitative aspects of thermochemical and physical fuel properties can be difficult. The ability of ML techniques to learn from previous datasets and procedures allows it to handle such problem [150].

A process of four steps is required to build a ML model described in **Figure 2.3** as follows:

- a. Building a dataset with appropriate samples
- b. Appropriate selection of descriptors and algorithms
- c. Using the database and descriptors to build the algorithm,
- d. Depending on the chosen descriptors, using the developed model to forecast the output values [151]

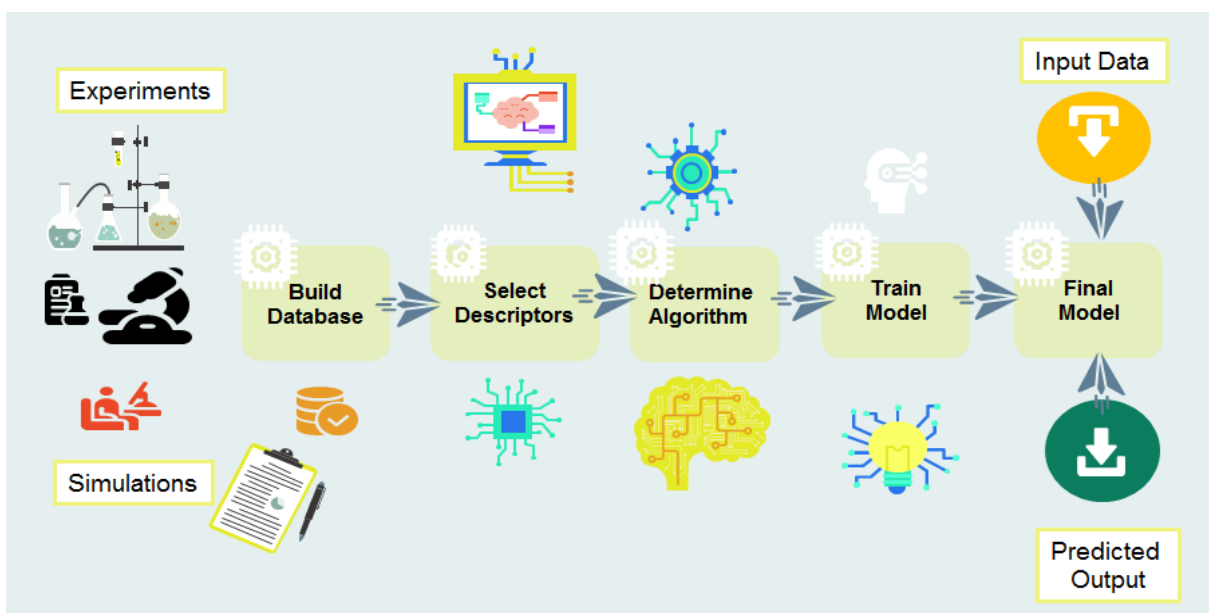


Figure 2.3: Schematic of a standard ML development model.

2.4.3 Historical Development of ML in Fuel Research

The application of ML algorithms, particularly ANNs, in fuel property prediction began in 1996. Keshavaraj et al. used ANNs to predict the fluid properties of siloxanes, employing descriptors like AET, Branching Patterns (BP), and Topological Connectivity (TC) to measure critical volume, temperature, and pressure [152]. In the same year, Hall et al. developed QSAR (Quantitative Structure-Activity Relationship) models using ANNs to predict boiling points and critical temperatures for datasets of 298 and 165 compounds, respectively, relying on the electro-topological state of atoms [153]. Subsequent advancements include:

- **2000:** McClelland et al. used a dataset of 420 molecular structures to predict vapor pressure with eight CNN descriptors, including electronic, hybrid, geometrical, and topological features [129].
- **2006:** Gao et al. implemented BPANN and multiple linear regression (MLR) models to develop QSPR (Quantitative Structure-Property Relationship) models for predicting melting and boiling points of polyamides [154].
- **2007:** Katritzky et al. built QSPR models using multi-layer perceptron algorithms to predict vapor pressures of 645 organic compounds [155].
- **2008:** Yong Pan et al. enhanced QSPR models with Support Vector Machines (SVMs), predicting flash points for 1,282 organic molecules classified into 57 functional groups [130].

- **Saldana et al.:** Combined multiple ML models to predict flash points and cetane numbers using SMILES (Simplified Molecular Input Line Entry System) formulas and functional group descriptors, demonstrating the integration of graph-based approaches in QSPR modelling [62].

These efforts highlight the increasing adoption of ML techniques in fuel research. The application of perceptron-based architectures, such as three-layer perceptron's with backpropagation algorithms, has further advanced the field. Basu and Henshaw used such architectures to predict adsorption isotherms of volatile organic compounds [131].

CHAPTER 3

SP MODELLING

The study aims to develop an advanced model to predict the SP of fuels containing oxygenated groups specifically ketones, alcohols, and ethers. The combinations of hydrocarbons with oxygenated compounds are nearly limitless, which presents a significant challenge for fuel studies. While Qasem's model focuses on ethers and alcohol in real fuels[25], this work introduces ketones to extend SP prediction using TPRF. This study addresses the gap that traditional models struggle with, while complex fuel blends reflect challenges to predict SP. Establishing an efficient methodology for predicting the SP of both new and current fuel blends is crucial. By enabling proactive soot reduction during fuel formulation, mixture ratio modifications according to anticipated sooting tendency could be possible. As a result, combustion engines will operate depending on environmentally friendly fuels, reducing the impact of soot emissions on the environment [156–158].

3.1 Methodology

Research on predicting SP has the ability to catalyze the development of atmospheric models aimed at minimizing the environmental impact of smoke emissions [159]. Model development was achieved using SP data from 89 experimental measurements. This large exhaustive experimental dataset (containing blends of TPRF and oxygenates) was used for model development, validation and testing.

The predictive model was developed using artificial neural networks (ANNs) as traditional regression methods have proven inadequate for accurately capturing the complex relationships between fuel chemistry and behavior [160].

Figure 3.1 visualizes the overall process for predicting SP containing the necessary stages, providing a roadmap for model development. Details explanation for selecting oxygenate is explored below:

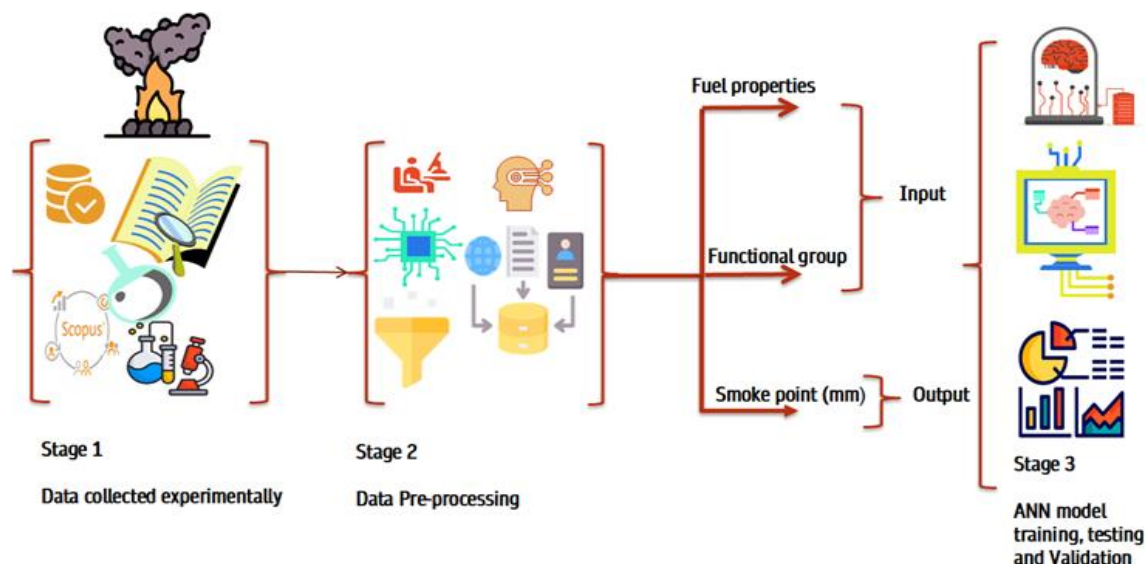


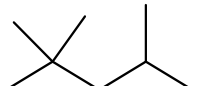
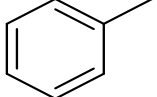
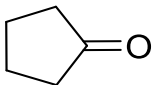
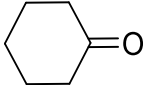
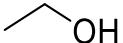
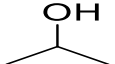
Figure 3.1: Flowchart for developing SP prediction model

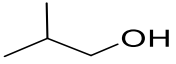
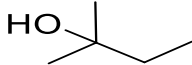
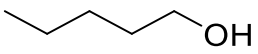
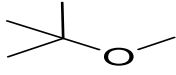
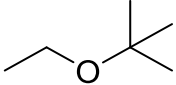
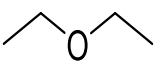
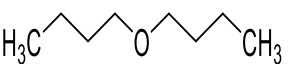
3.1.1 Selection of Oxygenated Compounds

In preparing the fuel blends for testing, the primary objective was to match the Research Octane Number (RON) and density of commercially available gasoline, ensuring that the surrogate blend accurately reflects key characteristics such as ignition behavior, vaporization properties, heat release rates, and combustion stability.

This approach ensures the surrogate is representative of the fuels commonly found in the consumer market, making the testing results applicable to real-world conditions. To achieve this, a variety of oxygenated compounds were selected, focusing on their ability to influence these combustion properties. These additives include Ethanol (EtOH), n-Pentanol (n-PeOH), Isopropanol (i-PrOH), Isobutanol (i-BuOH) 2-Methyl-2-Butanol (2MB2OL), Methyl tert-Butyl Ether (MTBE), Ethyl tert-Butyl Ether (ETBE), Diethyl Ether (DEE), Dibutyl Ether (DBE), Cyclohexanone (CHN) and Cyclopentanone (CPN). The properties of these oxygenates are summarized in **Table 3-1**.

Table 3-1: Physicochemical Properties of Selected Chemicals for SP Measurement

Chemicals	Structure	O w%	H/C	MW (g/mol)	Density (g/cm ³)	SP (mm)	Ref
n-Heptane		0	2.29	100.21	0.684	137	[161]
Iso-octane		0	2.25	114.23	0.692	49	[161]
Toluene		0	1.14	92.14	0.867	8.9	[161]
CPN		19.04	2.0	84.12	0.948	26.2	[31]
CHN		16.33	2.0	98.15	0.95	26.2	[31]
EtOH		34.73	3.0	46.07	0.789	-	
i-PrOH		26.64	2.67	60.1	0.785	-	

i-BuOH		21.59	2.75	74.12	0.802	41.7	[162]
2MB2OL		18.18	2.5	88.15	0.81	25.9	[31]
n-PeOH		18.18	2.4	88.15	0.814	27.7	[31]
MTBE		18.15	3.0	88.15	0.74	24.1	[31]
ETBE		15.66	2.8	102.18	0.744	-	
DEE		21.59	3.0	74.12	0.713	-	
DBE		12.31	2.25	130.23	0.769	-	

In this work, TPRF blends is the selected base fuel instead of more complex commercial gasoline, also to simplify the analysis. The chosen oxygenates were selected based on their availability, relevance to real-world applications, and prior research indicating their effectiveness in reducing soot[161]. The varying chemical structures of these oxygenates allowed for deeper exploration of how the C/H bond and oxygen atom position affect soot formation, providing valuable insights for cleaner combustion processes. TPRF surrogates were blended with 2%, 4%, 6%, 8%, 10%, 12% and 14% (by volume) of the oxygenated compounds.

3.1.2 Virtual Smoke Point VSP

The concept of the VSP was developed as an innovative approach to address challenges in measuring the SP of certain fuels, such as n-heptane and isooctane, which fall beyond the range of ASTM D1322 standard values.

Haas et al.[163] utilized a linear blending approach to overcome this limitation, based on the TSI to predict the SP for fuels with minimal sooting propensity.

In this study 23 fuel blends fall under VSP values, these mixtures contain PRF as a based fuel with oxygenated additives, and TPRF20-10 summarized in Table 3-2 These findings lead us to select TPRF80-20 as the base fuel for this work. The predictive model didn't include the VSP data collected through experimental measurements, providing further accuracy to prediction, and offering a deeper insight into the behavior of fuels at their SP.

Table 3-2: Initial investigation of the fuel blend

Base Fuel	Additives	Concentration
PRF20	n-PeOH	5%, 10%, 15%, 25%
PRF20	CPN	10%, 20%, 30%
PRF20	2MB2OL	10%, 20%, 30%
PRF20	DEE	10%, 20%, 30%
PRF20	MTBE	10%, 20%, 30%
PRF20	-	-
PRF30/40/50/60	-	-
TPRF20-10	-	-

3.2 Experimental Setup

The experimental setup for measuring the SP was carefully designed to ensure accuracy and consistency, following the ASTM D1322 Standard[164]. An SP lamp apparatus was used (see **Figure 3.2**), and oxygenates were incorporated into TPRF blends, based on volume percentages under the fuel preparation step where the blended fuel was stirred, and to ensure a homogenous blend, the mixtures were stirred using a magnetic stirrer at 550 RPM for 1 minute and 40 seconds, ensuring no phase separation or sedimentation occurred, then a cotton wick was soaked in the fuel sample for two minutes to ensure full saturation, followed by trimming the wick to a uniform length of 125 mm. This trimming was essential for maintaining consistency across tests. The wick was then positioned correctly within the candle apparatus, as misalignment could lead to uneven burning or flame distortion.



Figure 3.2: Apparatus setup for SP Measurement, ASTM D1322

The wick was fully saturated with 20 ml of fuel sample loaded to the candle, and inserted into the lamp apparatus, the wick guide ignited, and the chamber closed to preserve the flame see **Figure 3.3**:

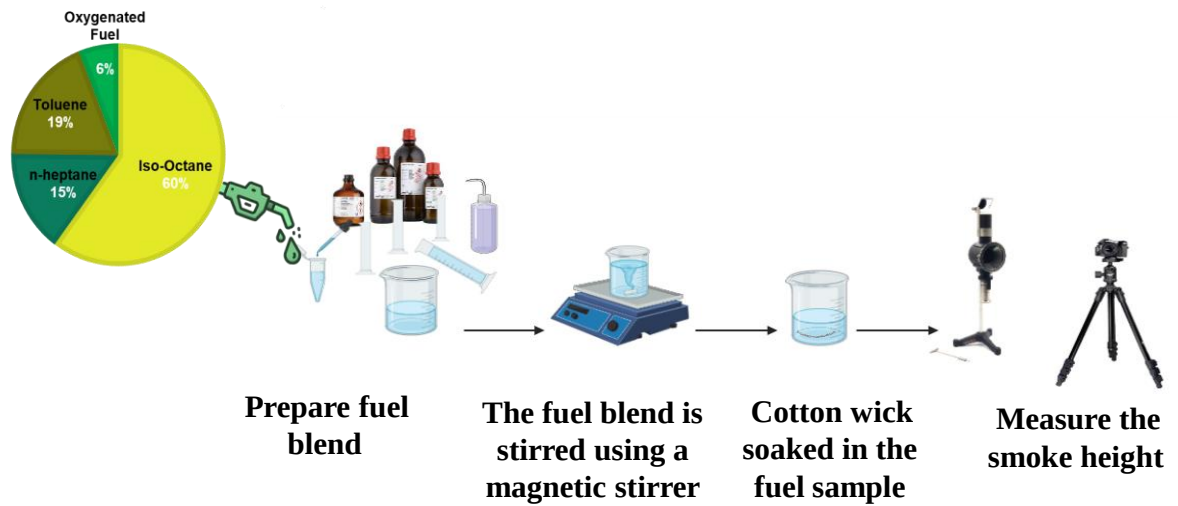


Figure 3.3: SP measurement steps

The initial flame was stabilized at 10 mm for five minutes, then the flame height gradually increased until a soot tail appeared, after which it was lowered until the soot disappeared, for 5 minutes. a paper was used to ensure there was no soot at this length marking the 1st SP reading. The flame height was recorded with a tolerance of ± 1 mm using a Nikon D5600 camera, and all of the equipment was thoroughly cleaned between tests using n-heptane and ionized water to avoid contamination[25].

Figure 3.4 represent the steps for reading SP for the TPRF80-20+Cyclopentanone6%. To ensure precision, the flame height was measured three times, and the average height was used. This method provides a reliable estimate of the fuel mixture's sooting propensity, as the SP is inversely proportional to soot production[27].

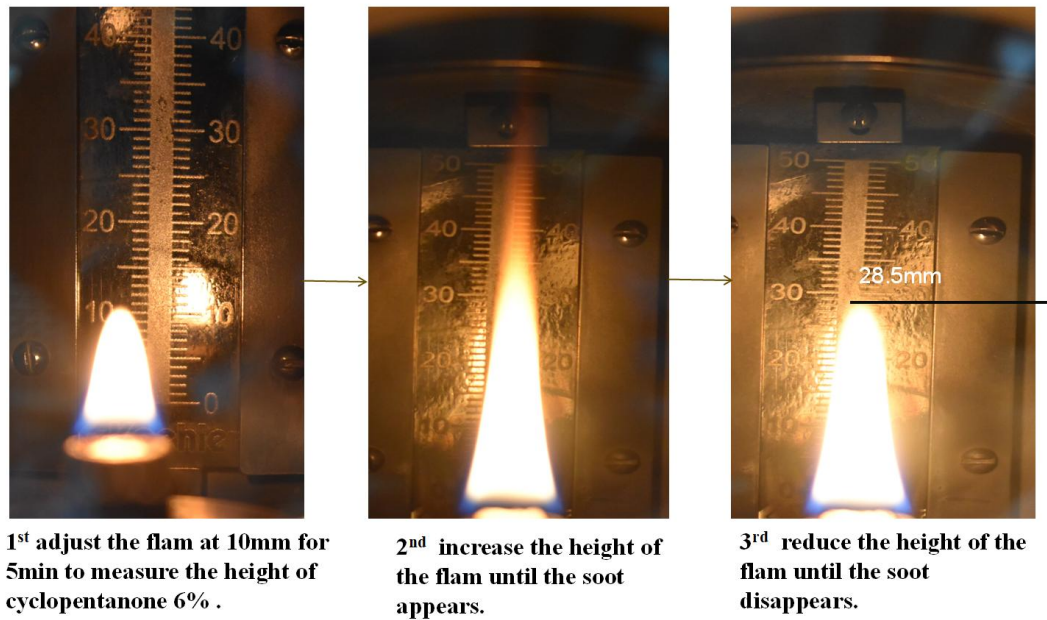


Figure 3.4: Steps for reading SP height for TPRF80-20 blended with cyclopentanone6%

3.3 Computational Setup

In this study, all models were implemented on a personal Dell Latitude 5530 laptop using Python 3.11/3.12 (64-bit) as a programming language due to the extensive library supported via it, as well as its adaptability in handling complex computations.

With data visualization and live code execution offered by Jupyter Notebook as an integrated development environment IDE. The setup, ranging from data collection to model development and results, will be discussed in the following sections. Additionally, the code used for implementing the models and analysis is available in the appendix for reference.

3.3.1 Dataset

This study used an experimental dataset comprising 89 fuel mixtures of TPRF with ketones, ethers and alcohols. The features of the dataset are 11 attributes, one of which was density, a widely studied property in ML models, and can be related to combustion performance. Other key features included BI, MW, and functional groups, all of which were determined based on the chemical structure of the compounds. The diversity of the data provided a robust foundation for ML models, enabling the prediction of SP for various fuel blends[117] [161].

A statistical analysis of the dataset yielded information about the features, which are shown in **Table 3-3**. The arithmetic mean (μ), maximum, minimum, and standard deviation (σ) of the characteristics are all included in **Table 3-3**. Such statistics comprehend the range and variability of the feature values for better understanding of the dataset. Primary, secondary, tertiary, and quaternary carbon atoms make up the paraffinic hydrocarbon functional groups denoted by the acronyms P.CH₃, P. CH₂, P.CH, and P.C. The functional groups denoting ketone, alcoholic, ether, and aromatic, functionalities are denoted by the terms CO, OH, O, and Arom.

Table 3-3: Statistically analyzing dataset features for SP prediction based on wt% alcohol OH%, ether O% , ketone CO%

	Dens ity	P.C H3 %	P.C H2 %	P.C H%	Aro m%	OH %	O %	P.C %	CO %	Mw	B.I	SP
No.	89	89	89	89	89	89	89	89	89	89	89	89
μ	0.73	47.8	20.4	6.71	16.9	0.73	0.5	6.35	0.61	104.	0.6	31
σ	0.01	3.50	2.94	0.81	4.18	1.12	0.8	0.76	1.38	2.9	0.05	4.5
Min	0.69	40.0	15.5	5.74	0	0	0	5.3	0	94.9	0.4	25.
Max	0.76	62.3	29.5	10.3	19.9	4.24	3.0	9.49	5.85	112.	0.61	49.5

In Regression methods, heatmaps provide insight investigation to understand the relationships between features and SP in model prediction. Pearsons's correlation (PC) is frequently utilized for such purpose **Figure 3.5**, illustrates how the eleven input features correlate with SP.

The values in the cross-section of the rows and columns labelled with the other feature indicate the degree of correlation between any two features. For instance, the correlation value of 0.52 between P.CH3 (wt%) and Mol.Wt (Mw) suggests that these features jointly influence SP. The aromatic content (Arom, wt%) and the BI show a positive correlation of 0.71, highlighting the role of both the positioning of aromatic functional groups and the presence of oxygenated groups in combustion characteristics. A higher BI indicates greater branching, which typically results in a lower SP due to its influence on molecular packing and combustion processes.

The BI, as introduced by Abdul Jameel et al. [87], is a method of measuring the degree of branching, considering both the size and the position of branches. A strong correlation (0.89) between the paraffinic content (P.C%) and P.CH3% further emphasizes the role of branching in influencing fuel properties. This suggests that increased branching leads to unique combustion behaviors, directly affecting the SP. While the negative correlation between Arom and SP (-0.92) means that higher Arom content reduces SP, experimental results provide a detail into the effects of TPRF blends which contain Arom group with oxygenated groups in providing higher SP. These relationships are crucial for optimizing fuel blends for better combustion.

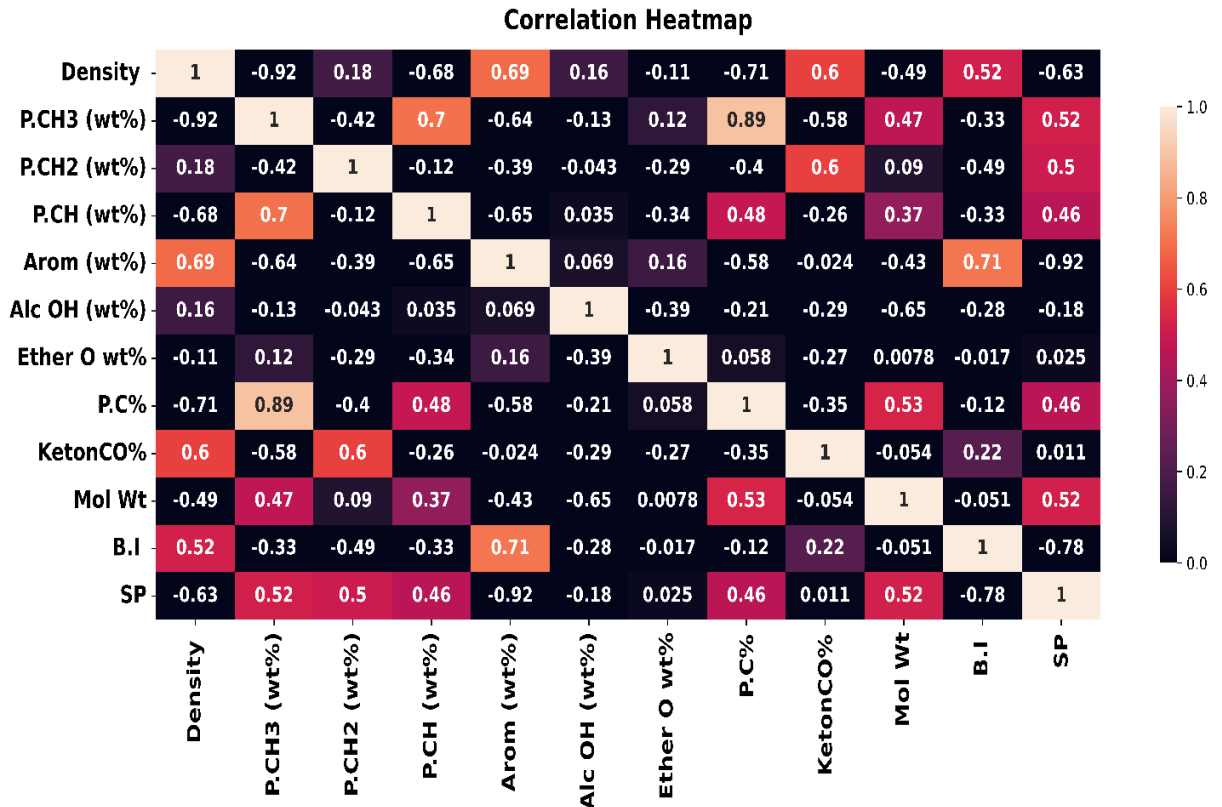


Figure 3.5: Heatmap of Pearson Correlation Coefficients to represent SP relationship with input features

3.3.2 Model Development

The capability of the ANN model to handle highly nonlinear, complex relationships is one of its strengths and has gained recognition and validation in the research [165]. The data preprocessing phase was conducted prior to model development to ensure consistent feature scaling and improve model performance. Standardization techniques were applied to the input features using the StandardScaler from the sci-kit-learn library, normalizing each feature to have a mean of 0 and a standard deviation of 1. This process was complemented by visual checks to validate correct scaling across features. After preprocessing, the dataset was randomly split into training (70%), validation (15%), and testing (15%) subsets, a structure intended to enhance model training fidelity and enable reliable performance evaluation. Once the data has been preprocessed, an ANN was designed, and its architecture optimized through hyperparameter tuning using Optuna. important hyperparameters such as neuron count, layer depth, learning rate, and dropout rates were adjusted to mitigate overfitting and improve model generalization. An exploration of regularization techniques followed, starting with L1 regularization to bolster stability, and progressing to L2 regularization to achieve improved parameter sparsity. Further adjustments included evaluating multiple activation functions such as: tanh, ReLU, and Leaky ReLU to determine the optimal configuration for the model. Model validation was performed using K-fold cross-validation, with mean squared error (MSE), mean absolute error (MAE), and regression coefficient (R^2) metrics providing a comprehensive assessment of the model's accuracy and generalization capabilities. This rigorous validation framework ensured a well-evaluated and robust ANN model[25,132,133,156,160,166–175].

CHAPTER 4

RESULTS

4.1 Experimental Results

Different oxygenated groups, such as alcohols, ketones, and ethers, exhibit varying sooting behaviors compared to pure hydrocarbons. For oxygenates, the breakdown products frequently include oxygen-containing compounds, such as formaldehyde (CH_2O) and carbon monoxide (CO), which do not contribute to soot formation [176]. As a result, oxygenates typically exhibit a lower tendency to produce soot. However, the extent to which a particular oxygenate generates soot is strongly influenced by the nature of the functional groups (FG) it contains[36], not only the percent of the oxygen content[177].

4.1.1 Impact of Alcohols

The variation in the SP values at the different percent of oxygenated alcohols is presented in **Figure 4.1**:

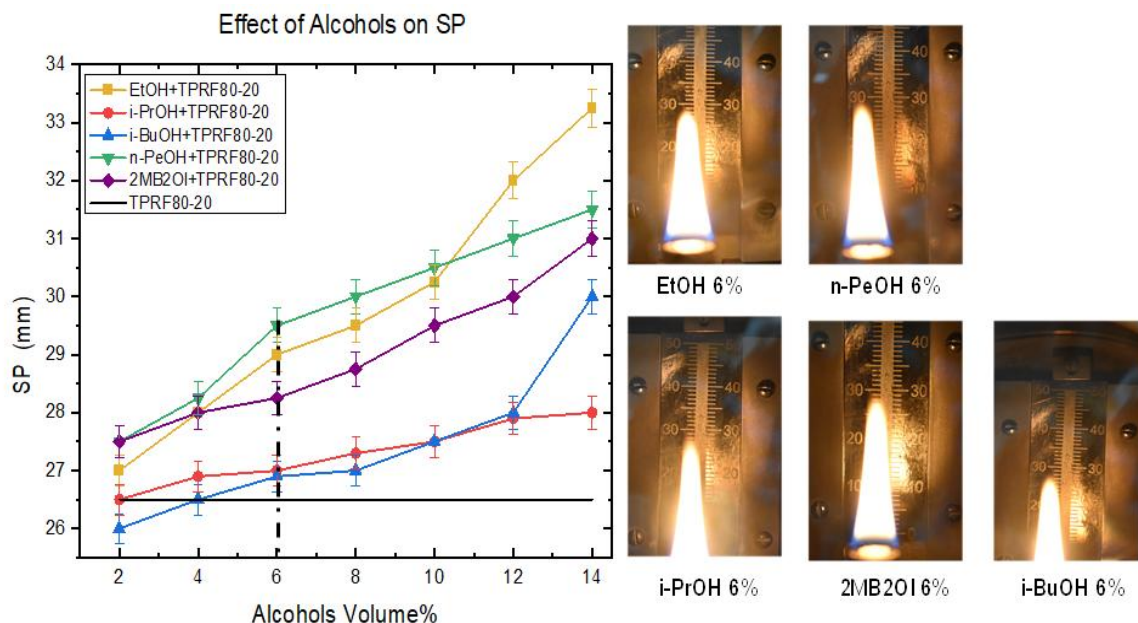


Figure 4.1: Impact of blending alcohols with TPRF 80-20 on SP

Straight-chain alcohols like EtOH and n-PeOH generally contribute to higher SP values due to their linear molecular structure, which promotes cleaner combustion and reduced soot formation similar to alkanes [31]. EtOH functions as a pure fuel, an oxygenate extender for petroleum-derived fuels, and also as an octane enhancer, reducing soot formation and improving overall combustion efficiency when blended with TPRF80-20 [90,178]. EtOH has a short chain with a high hydrogen-to-carbon (H/C) ratio, high O%, and low Mol Wt [179] . which results in higher SP compared to n-PeOH when blended with TPRF80-20 . While n-PeOH has a longer carbon chain, it shows a higher SP when blended with TPRF80-20 at concentrations below 10%, likely due to its thermal stability, which interrupts soot precursor formation.

However, at concentrations above 10% in the TPRF mixture, EtOH's SP surpasses that of n-PeOH, likely because ethanol's higher volatility and oxygen contribution make it more effective in reducing soot at elevated concentrations, promoting more complete combustion. EtOH is also known to have a shorter ignition delay compared to n-PeOH due to its high volatility, providing highly uniform mixing with air, minimizing fuel-rich zones, and hence potentially decreasing soot production.

Branched alcohols display unique sooting tendencies due to the impact of molecular branching on combustion behavior [31,87]. Introducing branched hydrocarbon structures, such as in alcohols like i-PrOH, i-BuOH, and 2MB2OL, regularly results in more complex carbon networks that are less likely to fully decompose, thereby aggregating into soot particles. This trend was also highlighted by Barrientos et al. [162], who found that branching often increases the sooting tendency, as steric hindrance from branching inhibits complete oxidation.

For 2MB2OL, with its additional methyl group, a highly branched structure, and a longer carbon chain but similar O% content as n-PeOH, SP values are higher than those of i-PrOH and i-BuOH. This is likely due to the methyl group's steric hindrance, which resists decomposition into soot precursors and delays soot formation. For i-PrOH and i-BuOH, the SP values reflect their combustion stability differences. At low concentrations (<10%) in TPRF blends, i-PrOH has a higher SP than i-BuOH, possibly due to its shorter carbon chain that provides high volatility and lower energy content, which leads to more controlled combustion.

At higher concentrations, however, i-BuOH exhibits a higher SP than i-PrOH, which allows for stable flame behavior at these levels by delaying soot formation. These branching effects are consistent with Barrientos et al.'s findings, which indicate that branched alcohols tend to have complex sooting behaviors due to their varied decomposition and oxidation rates, meaning they decompose into a variety of intermediate products that do not easily form PAHs, the building blocks for soot, which increase the SP and reduce PM and NO_x [31,162].

4.1.2 Impact of Ethers

Based on previous research [31,180], certain oxygenated functional groups have proven more effective than others at reducing soot and particulate emissions (PE). Ethers generally tend to perform better than alcohols [25] at soot reduction. Ethers exhibit a symmetrical molecular structure, which leads to a unique breakdown process during combustion: the molecule splits at the oxygen atom, producing small, non-sooting fragments. This fragmentation process results in higher SP compared to alcohols of similar molecular weight due to the reduced likelihood of forming soot precursors.

For ethers, the oxygen atom breaks the carbon chain, leading to smaller byproducts that do not easily form benzene rings reducing soot formation. McEnally et al. [181], highlighted that the molecular structure significantly influences the decomposition process, which explains why ethers produce smaller, less reactive fragments, contributing to reduced soot when used in TPRF fuel blends.

For instance, adding ethers like DBE to a blend of TPRF80-20 significantly increases the SP, thereby reducing the sooting tendency of the fuel mixture. DEE and ETBE also show similar positive effects, although to a slightly lesser degree than DBE.

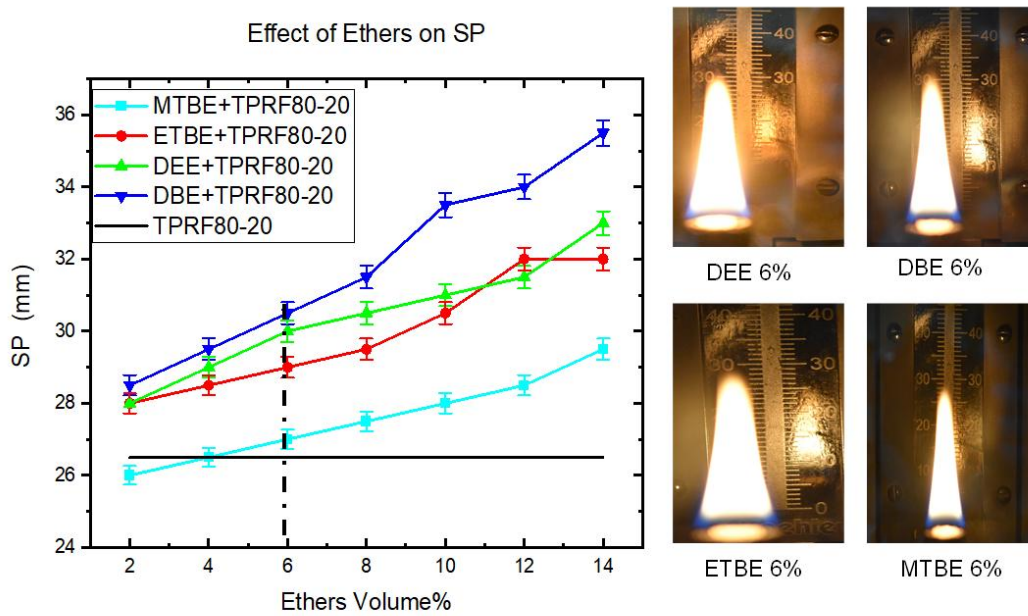


Figure 4.2: Impact of blending ethers with TPRF80-20 on SP

As shown in **Figure 4.2**, when blended with TPRF80-20 at 6%, DBE has the highest SP value among the ethers tested. Research by [177] analyzed PM emissions from a compression ignition engine under varying load conditions, comparing DBE's performance to alcohols for soot reduction. The study demonstrated that DBE achieved an 80% reduction in particulate emissions relative to diesel fuel, confirming its effectiveness in soot control. DBE's high SP can be attributed to its ether symmetric structure, which consistently fragments into non-aromatic, non-sooting byproducts.

In contrast, MTBE showed the lowest range of SP values among the tested ethers when blended with TPRF80-20. MTBE's structure, with a branched configuration, appears less effective in reducing soot precursors, leading to a lower SP than DBE. Looking at DEE across a volumetric concentration range from 2-8%, DEE exhibits similar characteristics to DBE, with a high SP value comparable to DBE at low-to-moderate concentrations. However, above an 8% concentration, the rate of increase in DEE's SP slows. At a blend ratio of TPRF80-20 with 12% DEE, the SP value was lower than that of ETBE, which started with an initial SP value similar to DEE at 2% but demonstrated a consistently lower rate of increase compared to DEE as concentration increased.

At higher concentrations (12-14%), ETBE and DEE show close SP values, indicating that the structural differences between them have a diminishing impact on SP at high blend ratios. On the other hand, MTBE demonstrates a steady, linear increase in SP across concentrations, indicating a consistent, albeit lower, effect on reducing soot formation relative to the other ethers.

4.1.3 Impact of Ketones

Ketones provide high energy conversion efficiency from chemical to thermal energy and possess strong knock resistance. Their lower cost also makes them commercially viable [182]. Investigations have observed that ketones enhanced stability during combustion comparable to alcohols such as EtOH [183]. Comparing both CPN and CHN as cyclic oxygenated additives based on **Figure 4.3** CHN shows high SP height starting from a volume concentration of 6%, while CPN has the highest value at a lower concentration of 2%. The outperforming behavior of CPN at low concentration is due to its ring strain, which tends to be reactive during combustion. The ring opening produces oxygen-rich fragments that burn cleanly without contributing to soot. CHN has a high H/C ratio [117], which assures a more complete combustion. More hydrogen atoms promote the formation of stable byproducts like water (H_2O) instead of carbon-based soot precursors.

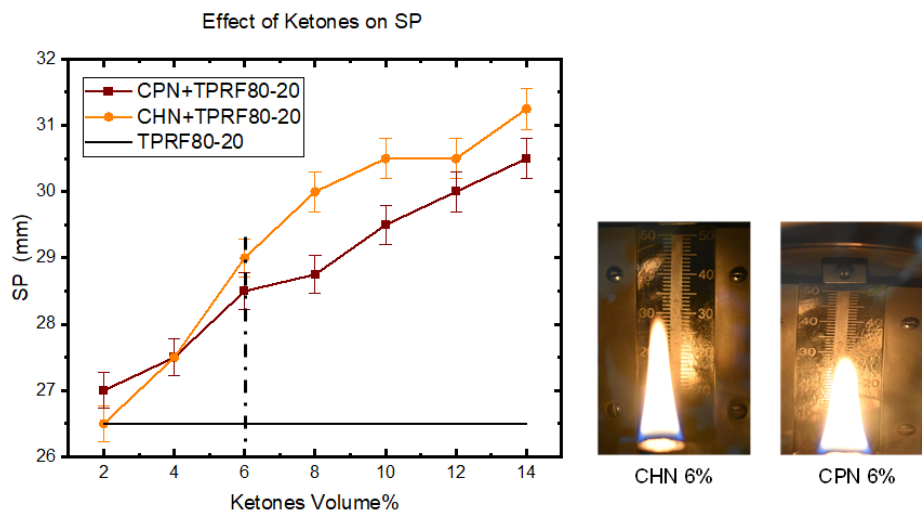


Figure 4.3: Impact of blending ketones with TPRF80-20 on SP

For ketones, such as CPN, the influence on SP is gradual, with noticeable improvements observed in the SP value compared to other oxygenates. Ketones do not participate in the four-center reactions like secondary alcohols, which makes them less prone to forming olefins that can contribute to soot formation. As a result, the use of ketones like CPN in TPRF fuel mixtures can effectively improve the SP, indicating a lower propensity for soot formation [181].

The differences in SP among the oxygenated fuel blends become more pronounced as the volume percentage of the oxygenated component in the blend increases. The broadening of SP measurements can be attributed to the influence of molecular weight (Mol Wt) and molecular structure of the oxygenated fuels at higher blend ratios. The impact of straight-chain versus branched-chain structures also underscores the role of the BI in determining smoke and soot formation characteristics. These insights emphasize the importance of studying each functional group independently to understand their distinct sooting behaviors comprehensively [157], [181], [184].

4.1.4 Threshold Sooting Index (TSI)

TSI values were measured for alcohols, ketones and ethers, based on **equation (9)**. The constant values a and b were calculated based on the experimental value of isooctane and toluene SP and TSI from [185]. The constant values for the SP apparatus are 4.637 and – 4.01 for a and b respectively. The reduction in the TSI of the base fuel comparable to the range of the additives blended with TPRF80-20 is presented in the **Figure 4.4**:

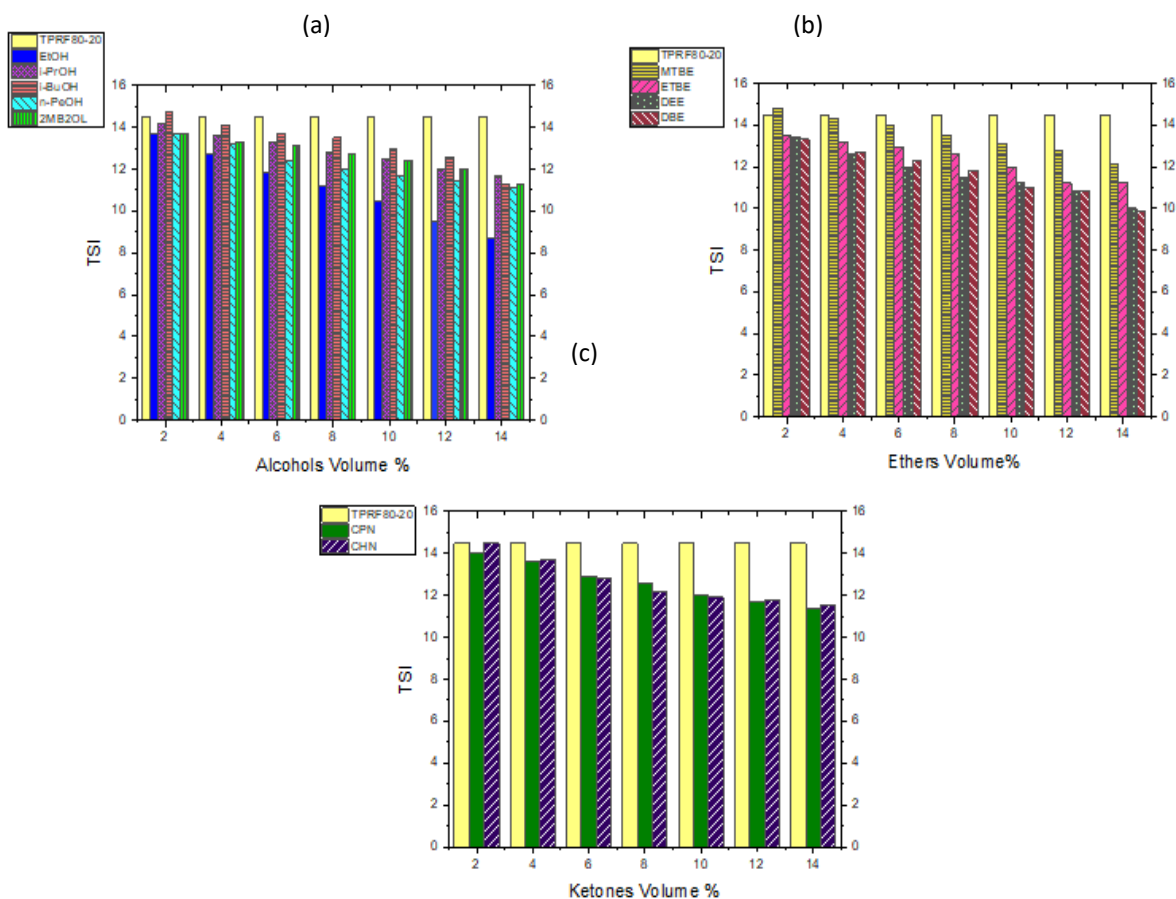


Figure 4.4: Impacts of adding alcohols(a), ethers(b), and ketones(c) with various percentages on TSI

The TSI value of the base fuel is 14.5 and adding oxygen content reduces this value significantly based on the volume percentage of the oxygenate. **Figure 4.4 (a)** illustrates the effects of TSI after adding various oxygenated groups to the base fuel TPRF80-20. Considering alcohols as an additive to the TPRF80-20, i-BuOH blend show a higher TSI values due to its lower SP, which has been affected by the branching, as we increase the i-BuOH the TSI decreases.

However, i-BuOH indicates higher sooting tendency comparable to other alcohols due to its higher TSI value across all percentages. The trend for other alcohols shows consistency in reducing the TSI value as the volume percentage increase, indicating a lower sooting tendency, among all alcohols increasing of EtOH percentage show a lowest TSI, at 14% blend of EtOH by volume, TSI show a value of 8.7. This shows that straight chain alcohols exhibit lower sooting tendencies comparable to the branching alcohols.

In **Figure 4.4 (b)**, ethers show a major decline of TSI, among all ethers, DBE blends with TPRF80-20 has the most effective results in reducing TSI to achieve 9.9 at 14% followed by the mixture of DEE reading TSI of 10 due to the effects of a high SP value and lower MW, while MTBE blends show a minimal TSI reduction at low concentration, after exceeding 10% addition, the rate of decreasing increase noticeably. ETBE exhibits an intermediate reduction in sooting tendency. Although the blends of CHN with TPRF80-20 have a higher SP value comparable to the CPN blend, the latter has a lower MW, which results in a lower TSI. Both ketones, as shown in **Figure 4.4 (c)** exhibit a consistent trend with a noticeable decrease in TSI as the volume percentage increases. **Figure 4.4** clearly indicates the role of adding oxygenate in reducing soot formation among all of the additives $\text{EtOH} > \text{DBE} > \text{DEE}$. Therefore, EtOH emerges as the most effective oxygenate in enhancing fuel properties due to its higher oxygen content, symmetry, and higher H/C ratio. Such results provide valuable information about the optimum selectin of fuel blends for cleaner combustion.

4.2 Computational Results:

The results from the implementation of deep learning and ML models for predicting SP provide insightful information about blending oxygenates with the base fuel. These results are explored below:

4.2.1 ANN model

Applying the ANN model discussed in the methodology, we selected the best parameters using grid search prioritizing accuracy and avoiding overfitting. These parameters were highlighted in **Table 4-1** after tuning.

Table 4-1: Significant characteristics of training parameters for ANN architecture

Parameters	Range	Optimum value
Regularization	L1/L2	L2
Number of Layers	1,2,3,4,5	4
Activation	‘tanh’ , ‘relu’	‘tanh’
Units/neurons	10,20,30,40,50,60,70,80,90,100 ,120,130,140	30, 70, 50, 1
Dropout	0.1,0.2,0.3	0.3
Learning rate	0.001, 0.002, 0.01, 0.02, 0.03, 0.1	0.001
Batch_Size	8,32, 64, 128, 256, 512	32
Epochs	300 <x <1500	600

Table 4-2 compares the predicted and actual SP values for the data set, provides concrete and quantitative evidence of the model's accuracy. This demonstrates the model's effectiveness in accurately predicting SP across a broad spectrum of fuel mixtures.

Table 4-2: Model performance metrics on TPRF80-20 blends

Model Performance	Training	Testing	Validation	All data
R^2	0.96	0.97	0.92	0.96
RMSE	0.18	0.23	0.2	0.19
MAE	0.13	0.17	0.14	0.14

ANN model's performance in predicting SP is represented in **Figure 4.5** and **Figure 4.6**. These cross graphs capture the influence of important factors such as molecular structure and the oxygen contents that affect SP. The hyperparameter used above helped enhance model sparsity, focusing on the most significant features and improving generalizability.

Figure 4.5 demonstrates the model's fit on the training dataset, achieving an RMSE of 0.18 and R^2 of 0.96, along with an MAE of 0.13. These metrics indicate that the model effectively learned the nonlinear relationships between input features and SP values, with predicted values closely matching the actual ones. Also, the figure represents the test dataset, the model also performed well, with an R^2 of 0.97, an RMSE of 0.23, and an MAE of 0.17. This strong performance on the test data suggests that the ANN model generalizes effectively to unseen data, with minimal overfitting.

Furthermore, the high R^2 values and lower error metrics suggested that ANN is a reliable tool for modelling the SP of complex fuel mixtures.

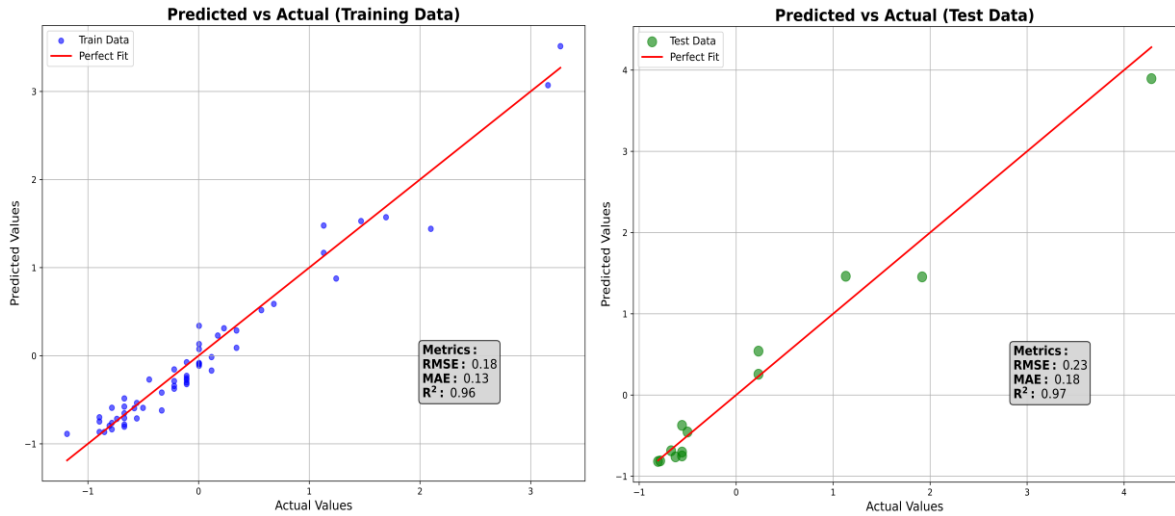


Figure 4.5: Comparison of experimental and predicted SP values for training and testing data using Ann model

Figure 4.6 illustrate the performance of ANN for both validation and entire data. The RMSE for all data is 0.19, while for the validation dataset, it stands at 0.20. This indicates that the model's average prediction error is low, with only a slight increase for the validation data. The MAE is consistent at 0.14 for both datasets, demonstrating stable prediction accuracy across all data points. Moreover, the R^2 score is 0.96 for all data and 0.92 for the validation set, meaning the model can explain 96% and 92% of the variance in each respective dataset. These results show that the model performs well, with strong accuracy and minimal prediction error. The slight differences in RMSE and R^2 scores between all data and validation data are typical and reflect a modest reduction in performance when transitioning to validation, which is a common sign of mild overfitting.

Overall, the model successfully captures the underlying relationships in the data and generalizes effectively, maintaining reliable accuracy even for previously unseen data.

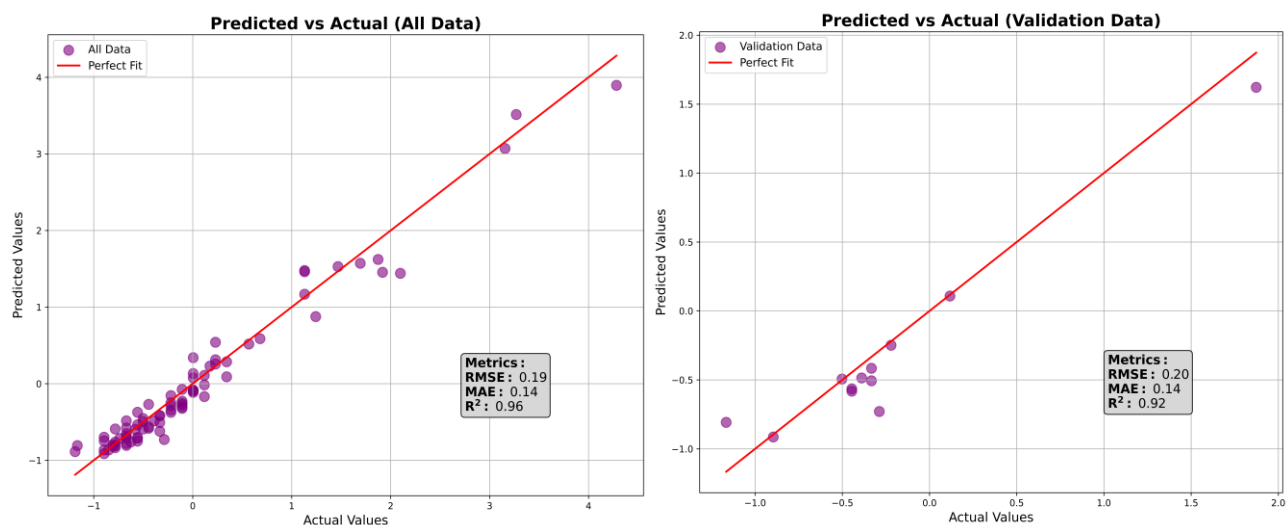


Figure 4.6: Plotting actual SP vs predicted SP for all and validation data using ANN model.

Eventually, the complex mixture with multi-functional groups and branched structures faced discrepancy while predicting due to the increment of various features interaction. Additionally, two fuel blends with identical SP will never give similar functional group distribution due to the unique nature of each mixture. Such obstacles add to the model's challenge. Incorporating a large and diverse data set will improve the accuracy of the predicted model. The selection of functional groups as a feature allows the model to quantify the fuel constituent component for more accurate and reliable prediction.

4.3 ML Models

To further investigate the complex and non-linear relationship between features, six ML models were used to predict SP these models include: extra tree regression (ETR), extreme gradient boosting regression (XGBR), Light gradient boosting regression (LightGBR), Nu- support vector regression (Nu-SVR), support vector machine (SVM), and random forest (RF). All the models were evaluated based on accuracy and generalization through the implementation of grid search for hyperparameter tuning. Detailed information about these models is provided below. The code used for implementing the models and analysis is available in the appendix for reference.

4.3.1 ETR Model

The ensemble learning method known as ETR [186] blends the ideas of random forests and decision trees. The ETR method guarantees decision resilience and minimizes overfitting. Trees by means of randomization. ETR does this by adding unpredictability to data sampling and feature sampling, two crucial steps in the tree-building process. ETR uses bootstrap sampling, which selects sections of the original data at random with replacement, as opposed to using the full training dataset for every tree. This lowers the possibility of overfitting by producing distinct training datasets for every tree. Furthermore, at each split, ETR randomly selects a subset of features rather than taking into account all of the features by performing random feature subsampling. As a result, the trees become more diverse.

The final ensemble forecast is created by combining the predictions of the various trees, usually by averaging them, producing a prediction that is more reliable and accurate [186].

4.3.2 XGBR Model

For regression tasks, XGBR (Extreme gradient boosting regression) [191] is another popular ensemble learning approach. An ensemble of weak regression models, or decision trees, is progressively constructed by fitting them to remnants from the prior model. The objective function, which is often an expression that computes the errors made by the earlier models, is the main focus of each succeeding model.

XGBR uses regularization approaches to penalize big weights or complicated tree structures in order to assess the model's complexity and avoid overfitting. The approach determines the ideal weights for each weak model using a gradient-based optimization algorithm, and a robust predictive model is produced by carefully combining these weak learners with the optimization of the objective function.

4.3.3 LightGBM Model

The implementation of Gradient Boosting Decision Tree (GBDT) models, such as XGBR, is computationally expensive because it involves scanning all data instances for each feature. To solve the limitations of efficiency and scalability on high-dimensional and large datasets that XGBR struggles with, researchers invented LightGBM.

LightGBM builds an ensemble of weak learners, typically decision trees, and uses a leaf-wise growth strategy, which splits the leaf with the maximum loss reduction rather than growing level-wise. This approach reduces training loss more effectively and often produces a better-fitting model compared to traditional level-wise strategies. LightGBM also employs advanced techniques like histogram-based learning, Gradient-based One-Side Sampling (GOSS), and Exclusive Feature Bundling (EFB) to enhance training speed and reduce complexity. By reducing the computational burden while preserving model accuracy, GOSS focuses on instances with larger gradients that contribute more to the model's learning process. It maintains data distribution by retaining all high-gradient instances and sampling low-gradient instances with a weighted adjustment. EFB merges mutually exclusive features—those rarely active simultaneously—into a single feature bundle, optimized through graph-based methods to reduce the feature space without significant loss of information, dramatically decreasing memory usage and computation time [192].

Eventually, LightGBM incorporates regularization techniques, making it a robust and scalable tool. The combination of speed, efficiency, and accuracy makes LightGBM particularly suitable for large-scale datasets and high-dimensional features, addressing the needs of academic and technical contexts.

4.3.4 Nu-SVR Model

Nu-SVR a supervised learning methods represented by [187] is a combination of Nu parameter with support vector regression SVR for regression tasks. Regression analysis carried out by support vector machine SVM utilized by SVR.

It attempts to keep the error within a predetermined range while identifying the hyperplane that best fits the data. Nevertheless, Nu-SVR improves this procedure by incorporating a "nu" parameter that regulates the error tolerance and the quantity of support vectors. A set of fundamental functions is used to estimate the regression function when working with non-linear kernels, producing a modified version. C is the penalty parameter linked to the error term in the Scikit-learn implementation of the SVR [188], regulating the trade-off between model complexity as well as mistake tolerance. Furthermore, the highest limit for the percentage of training errors is indicated by the parameter ϵ . The optimization problem and the description of the margin provided by the Nu parameter are where SVR and Nu-SVR diverge. The trade-off between the complexity of the model and the number of support vectors is controlled by the value of Nu. A more complex model may result from a smaller Nu value, which permits more support vectors. In contrast, a simpler model may be produced by limiting the number of support vectors with a higher Nu value.

4.3.5 Support Vectors Machine

Support Vector Machines (SVMs) are highly effective algorithms, particularly suitable for small and intricate datasets. These models prioritize specific observations, known as support vectors, which carry significant information, thereby reducing the impact of outliers on the overall performance. The approach encompasses both linear techniques, such as Linear Support Vector Regression (LinSVR), and nonlinear methods employing kernel functions, like SVR and Nu-SVR.

Support Vector Regression (SVR), a popular choice for tackling regression challenges, seeks to derive a function that best represents the data. Unlike conventional linear regression, SVR introduces an ε -insensitive zone, where errors within this margin between actual and predicted values are not penalized, ensuring robustness to minor deviations[189].

4.3.6 Random Forest

Multiple decision trees are constructed and averaged using the Random Forest (RF) algorithm, which is an ensemble approach. This method was created to overcome the shortcomings of basic Decision Trees (DT), such overfitting, by constructing a "forest" of trees in which individual trees' incorrect predictions are offset by those of other trees. RF is a reliable method for both classification and regression tasks, produces more consistent results, and is less susceptible to noisy data.

In contrast to a single decision tree model, Random Forest shows significant efficiency in this study and offers a baseline for model comparison (Sahin 2020). According to Segal et al. (2004), sample bootstrapping is a technique that, instead of learning from the complete dataset, creates a single decision tree from a sample.

Breiman (2001) first officially introduced the random forests approach, which uses bootstrap sampling to choose randomly a forest of decision trees from the previous training set and then averages the trees' output. The model's training complexity is expressed as $\mathcal{O}(n \log(n) \cdot d \cdot k)$, where n is the number of training instances, d is the data's dimensionality, and k (estimators) is the ensemble's number of decision trees, which in this work is fixed at 200.

A seed sample is taken from the training set to create the Random Forest model, and a decision tree is constructed with the max depth factor set to regulate the tree's depth. At every division in the assembly process, a portion of features is picked at arbitrary, and the node is split using the best feature. While the min samples leaf option stipulates that each leaf node must include at least two samples, the min samples split parameter guarantees that a node must have at least five samples in order to be further divided. To create the ensemble of decision trees, this procedure is repeated for the specified number of estimators[190].

4.4 SHAP Analysis

An effective tool for comprehending how predictions are made by ML models. Shapley additive exPlanations SHAP was first presented by Lundberg and Lee (2017) and gives input features priority. Demonstrating their respective significance in the results of predictions. Shapeley numbers, which quantify the impact of each feature on predictions, are computed by SHAP. It has gained popularity due to its simply explained characteristics and use of cooperative game theory. SHAP calculates the contribution and impact of each feature on the prediction of the final model using an additive feature attribution method [193].

4.5 ML Models Performance

The dataset utilized in this study where six ML models implemented to evaluate their suitability for prediction consisted of 89 experimental TPRF blends combined with oxygenated ketones, ethers, and alcohols. Eleven key features were identified for predicting the SP, encompassing properties such as density, MW, and BI. To ensure uniformity and prevent features with larger magnitudes from disproportionately influencing the model, all features were normalized using the StandardScaler from the sklearn.preprocessing module. This standardization transformed each feature to have a mean of 0 and a standard deviation of 1, ensuring that all features contributed equally during model training, effectively mitigating scale-related biases and enhancing the predictive accuracy of the models. Following normalization, the dataset was divided into training (80%) and testing (20%) subsets, with a fixed random state ensuring reproducibility across multiple experiments. The training data was utilized for model fitting and hyperparameter tuning, while the testing set was reserved for evaluating model performance on unseen data. Hyperparameter optimization, a critical step in enhancing model accuracy and generalizability, was conducted using GridSearchCV. This methodical approach involved systematically iterating over various parameter combinations, leveraging k-fold cross-validation ($k=3$) to evaluate model performance for each configuration. Grid search not only provided an efficient mechanism to identify the ideal hyperparameters but also minimized overfitting by ensuring robust evaluation across data folds. Compared to manual tuning, this approach offered a data-driven optimization strategy that was both time-efficient and reliable.

The parameter ranges explored during the grid search, along with the corresponding best-fit values for each model, are summarized in **Table 4-3**. Optimizing these parameters led to significant improvements in the models' predictive accuracy, generalizability, and alignment with the underlying data patterns. This rigorous optimization process ensured that the ML models were tailored to the specific characteristics of the dataset, enhancing their utility in predicting the SP of diverse TPRF blends.

Table 4-3: Configuration of hyperparameter tuning for ML models and best-fit values

Model	Parameters	Range	Best Fit
RF	n_estimators	[100, 200, 500, 1000]	200
	max_depth	[5, 10, 20]	20
	min_samples_split	[2, 5, 10]	5
	min_samples_leaf	[1, 2, 4]	2
SVR	Kernels	Polynomial, Linear, RBF	Linear
	Gamma	Scale, auto	scale
	C (Regularization)	[0.1, 1, 10, 100]	100
LGBM	learning rate	[0.01, 0.1, 0.2]	0.01
	n-estimators	[100, 200, 500]	500
	num_leaves	[31, 50, 100]	31
	max_depth	[5, 10, 15]	5

	min_data_in_leaf	[10, 20, 50]	10
XGBR	learning_rate	[0.01, 0.1, 0.2]	0.2
	n_estimators	[100, 200, 500]	100
	max_depth	[3, 5, 7]	3
	subsample	[0.7, 0.8, 1]	0.7
	colsample_bytree	[0.7, 0.8, 1]	0.7
ETR Model	n_estimators	[100, 200, 500, 1000]	500
	max_depth	[5, 10, 20, None]	20
	min_samples_split	[2, 5, 10]	5
	min_samples_leaf	[1, 2, 4]	1
Nu SVR	Kernels	Polynomial, Linear, RBF	Linear
	Gamma	Scaled , auto ,	Scale
	Nu (Support Vectors)	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]	0.5
	C (Regularization)	[0.1, 0.2, 0.3, 0.4, 0.5, 1, 1.1, 1.2, 1.4, 1.5, 1.6, 1.7]	1

4.6 ML Models Results:

In this section the results of the six ML model will be explored in detail, starting with ETR, then XGBoost, LightGBR, Nu-SVR, SVM and finally RF. Followed by a comparison between all of models using Taylor diagram and insight into feature importance and SHAP analysis.

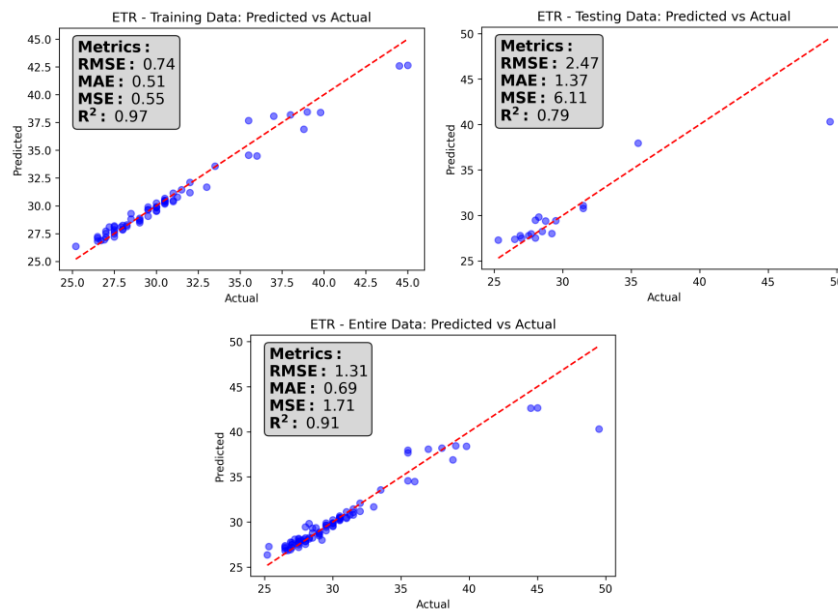


Figure 4.7: Predicted vs Actual values for training, testing, and entire dataset for ETR model

The performance of ETR model based on **Figure 4.7** for the training dataset demonstrated strong predictive capabilities, with an R^2 value of 0.97, indicating that 97% of the variance in the target variable was explained by the model. Additionally, the model achieved a low MAE of 0.51 and RMSE of 0.74, reflecting its ability to closely align predictions with actual values in the training data.

However, a significant decline in performance was observed on the testing dataset, where the R^2 dropped to 0.79, and the error metrics increased substantially, with an MAE of 1.37 and RMSE of 2.47. This suggests that the model overfitted to the training data, capturing noise rather than generalizable patterns. The scatter plot further highlights this overfitting, with several points deviating noticeably from the diagonal line, particularly for higher SP values. These results underscore the model's inability to generalize effectively to unseen data, reflecting one of the limitations of this model as it is less effective on small datasets. Emphasizing the need for mitigating overfitting, via expanding dataset.

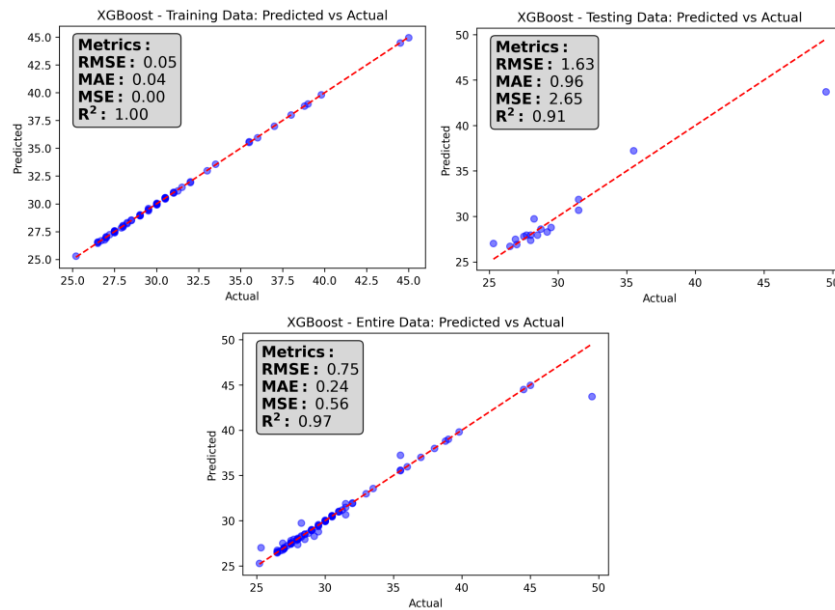


Figure 4.8: Predicted vs Actual values for training, testing, and entire dataset for XGBoost model

The performance of XGBoost model based on **Figure 4.8** for the training dataset demonstrated perfect predictive capabilities, with an R^2 value of 1.00, reflecting that 100% of the variance in the target variable was explained by the model. Additionally, the model achieved a minimal prediction error with MAE of 0.04 and RMSE of 0.05, suggesting the model closely matches the training data. However, a significant decline in performance was observed on the testing dataset, where the R^2 dropped to 0.91, and the error metrics increased substantially, with an MAE of 0.96 and RMSE of 1.63. This suggests that the model overfitted to the training data, the gap between training and testing metrics suggests the model has memorized patterns specific to the training data rather than learning generalized relationships.

The scatter plot further highlights this overfitting, with several points deviating noticeably from the diagonal line, particularly for higher SP values. These results underscore the model's inability to generalize effectively to unseen data, and to its ability to capture complex patterns through ensemble learning. reflecting one of the limitations of this model as it is complex to tune. Emphasizing the need for mitigating overfitting, via expanding dataset.

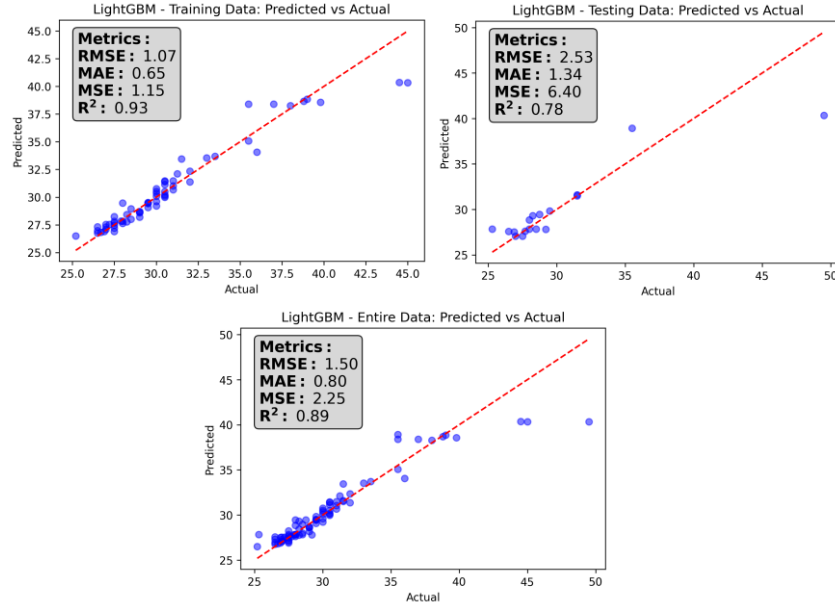


Figure 4.9: Predicted vs Actual values for training, testing, and entire dataset for LightGBM model

The performance of the LightGBM model based on **Figure 4.9** for the training dataset demonstrated satisfactory predictive capabilities, with an R^2 value of 0.93, indicating that 93% of the variance in the target variable was explained by the model. The model achieved a reasonable MAE of 0.65 and RMSE of 1.07, reflecting its ability to capture key patterns in the training data. However, a significant decline in performance was observed on the testing dataset, where the R^2 dropped to 0.78, and the error metrics increased substantially, with an MAE of 1.34 and RMSE of 2.53.

This decline suggests that the model struggled to generalize to unseen data, likely capturing specific patterns in the training data rather than learning generalized relationships. The scatter plot further supports this observation, with noticeable deviations from the diagonal line, particularly for higher SP values.

These results highlight the limitations of the LightGBM model in handling small datasets effectively, emphasizing the need for addressing overfitting by expanding the dataset or refining the feature engineering process.

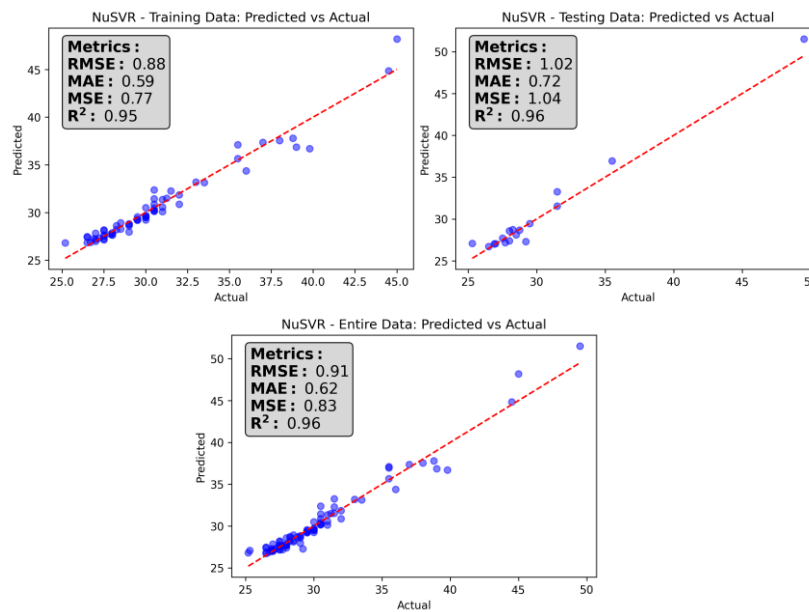


Figure 4.10: Predicted vs Actual values for training, testing, and the entire dataset for Nu-SVR model

The performance of the Nu-SVR model based on **Figure 4.10** for the training dataset demonstrated strong predictive capabilities, with an R^2 value of 0.95, indicating that 95% of the variance in the target variable was explained by the model. Additionally, the model achieved a low MAE of 0.59 and RMSE of 0.88, reflecting its ability to closely align predictions with actual values in the training data.

On the testing dataset, the NuSVR model maintained robust performance, with a slight decline in metrics; it achieved an R^2 value of 0.96, RMSE of 1.02, and MAE of 0.72. The minor gap between training and testing metrics suggests the model generalizes well to unseen data. The scatter plot highlights this strong generalization, with most points closely aligned along the diagonal line, indicating consistent predictive accuracy across the dataset. These results underscore the Nu-SVR model's ability to balance accuracy and generalizability, making it a reliable approach for predicting SP values in the dataset.

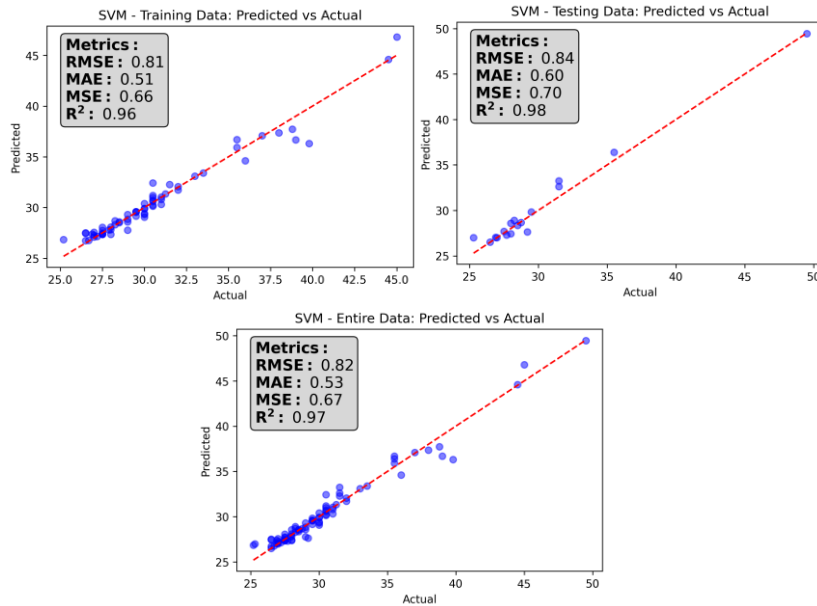


Figure 4.11: Predicted vs Actual values for training, testing, and entire dataset for SVM model

The performance of the SVM model based on **Figure 4.11** for the training dataset demonstrated strong predictive capabilities, with an R^2 value of 0.96, indicating that 96% of the variance in the target variable was explained by the model.

Additionally, the model achieved a low MAE of 0.51 and RMSE of 0.81, showcasing its ability to closely align predictions with actual values in the training data.

On the testing dataset, the SVM model maintained excellent performance, achieving an R^2 value of 0.98, an RMSE of 0.84, and an MAE of 0.60. The minimal gap between training and testing metrics demonstrates the model's robust generalization capabilities, effectively avoiding overfitting. The scatter plot further validates these observations, with points closely aligning along the diagonal line across both datasets, reflecting consistent predictive accuracy. These results highlight the SVM model's reliability for SP prediction, emphasizing its ability to generalize well while maintaining low prediction error.

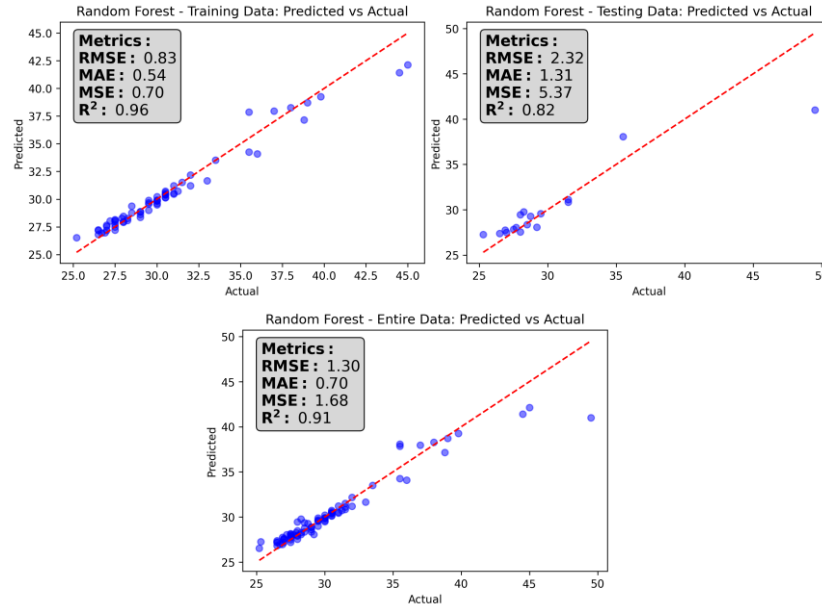


Figure 4.12: Predicted vs Actual Values for Training, Testing, and Entire Dataset for RF model

The performance of the Random Forest model based on **Figure 4.12** for the training dataset demonstrated strong predictive capabilities, with an R^2 value of 0.96, indicating that 96% of the variance in the target variable was explained by the model. The model also achieved a low MAE of 0.54 and RMSE of 0.83, suggesting its ability to align well with the training data. However, the testing dataset revealed a significant decline in performance, with the R^2 dropping to 0.82, RMSE increasing to 2.32, and MAE rising to 1.31. This considerable gap between the training and testing metrics indicates that the Random Forest model struggled to generalize effectively to unseen data, likely overfitting to the training set. The scatter plot supports this observation, showing noticeable deviations from the diagonal line for higher SP values in the testing data.

These results reflect one of the key limitations of the Random Forest model, emphasizing its sensitivity to overfitting, particularly when working with smaller datasets. This underscores the need for strategies to enhance generalization, such as increasing dataset diversity or exploring ensemble techniques to reduce variability.

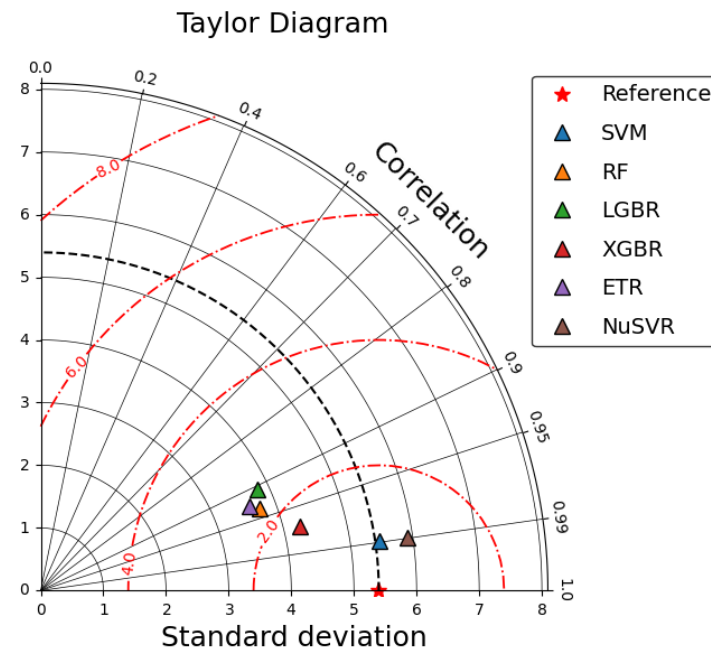


Figure 4.13: Taylor Diagram for Performance Comparison of All Models

The Taylor diagram compares the performance of several regression models, including ETR, XGBR, LightGBR, Nu-SVR, SVM and RF based on **Figure 4.13**. It visually summarizes three key metrics for each model relative to a reference (denoted by the red star): standard deviation, correlation coefficient, and centered RMSE. These statistics are crucial for evaluating how well a model represents the dataset.

A higher correlation coefficient indicates better alignment with the reference data, with a value close to 1.0 being ideal. The diagram shows that most models achieve a correlation coefficient between 0.8 and 0.9, demonstrating reasonably good predictive performance. However, models like SVM and Nu-SVR show a better fit compared to others, while the remaining models deviate more from the reference, indicating issues like overestimation or underestimation of the data's variability.

For RMSE, lower values signify that the model predictions are closer on average to the reference data. In this diagram, models like Nu-SVR and XGBR are positioned closer to the reference point, indicating lower RMSE and better prediction accuracy compared to the other models.

4.7 Feature Importance and SHAP analysis for ML models

Based on

Figure 4.14, the physical properties show a strong influence on prediction, density and MW are the most important factors in forecasting the SP, followed by aromatic content which is supported by persons correlation that indicates that the aromatic characteristics such as the double bond has highly negative effects on the SP value and P.CH content existence suggest the importance of H/C ratio in predicting sooting tendency of the fuel, according to the feature importance plot for the SVM model. For the ketones the important feature analysis requires more data to accurately show the effects of the ketones group while alcohols and ethers have more data in the model it shows significant importance in predicting SP.

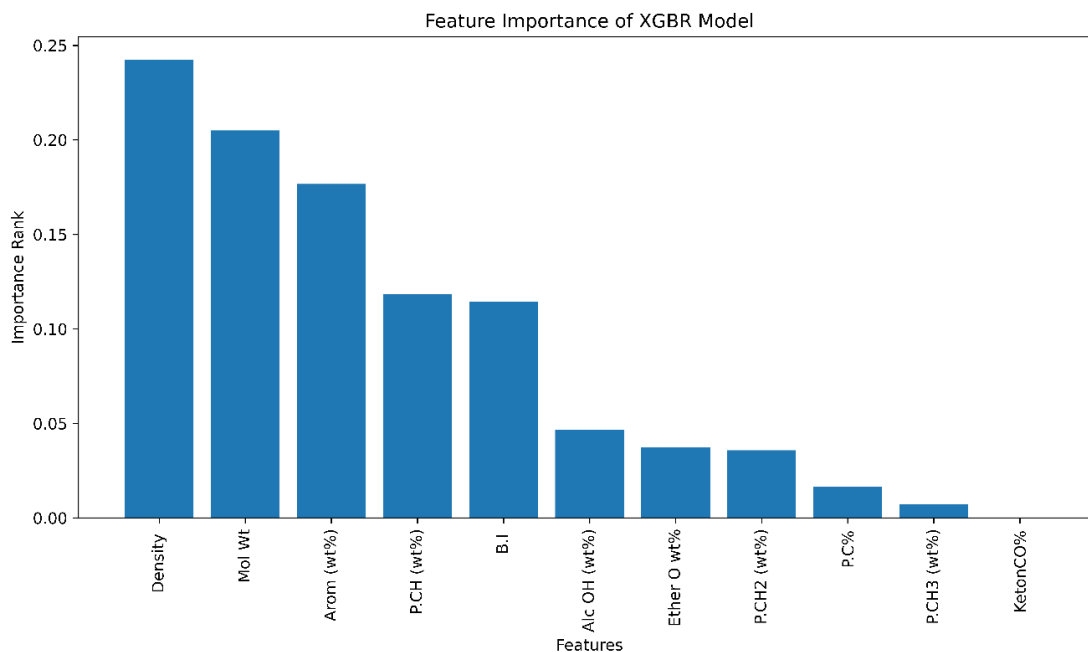


Figure 4.14: Permutation Method Illustration of Feature Importance on the SVM Model.

Also, BI's relation to the aromatics clearly appears in the plot highlighting the effects of branching on the SP as in the literature and this work. These findings highlight how important structural and compositional characteristics are in predicting the SP, provide insights into developing cleaner-burning fuels and contribute to a deeper understanding of the interaction of the molecular level in influencing fuel soot formation during combustion.

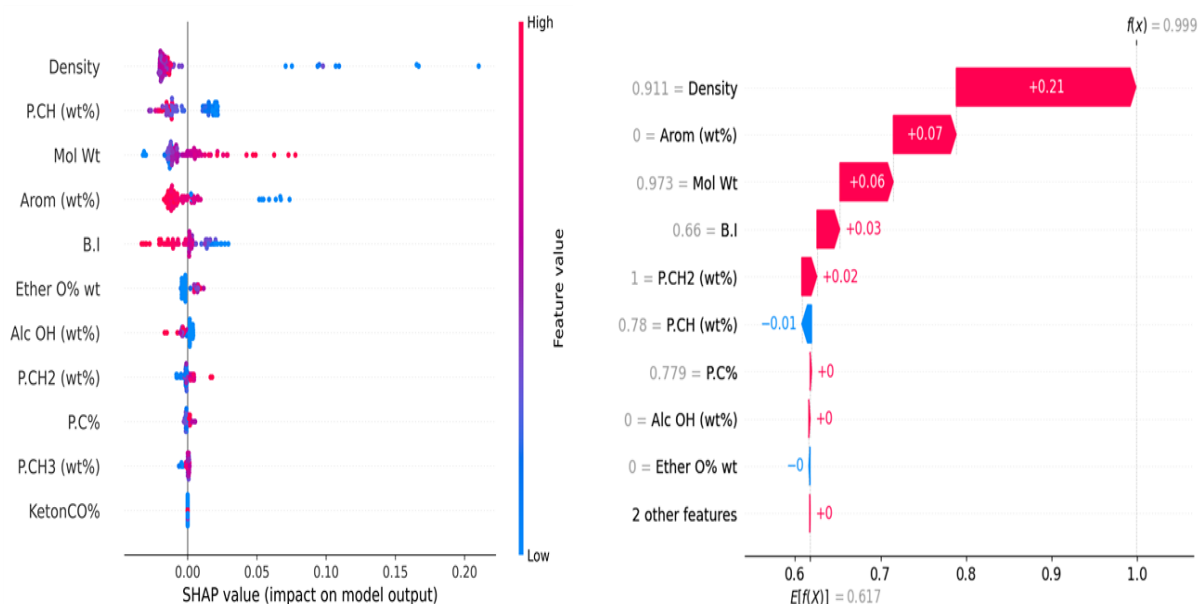


Figure 4.15: SHAP Summary Plot for Feature Impact on SP based on SVM Model

The SHAP as in **Figure 4.15** waterfall plot provides an in-depth explanation of how each feature contributes to the model output. It begins with the base value or expected value ($E[f(X)] = 0.617$), which is the average prediction of the model across the entire dataset. The bars in the plot represent the individual contributions of each feature to the final predicted value. The length and direction (positive or negative) of each bar indicates how much a feature increases or decreases the prediction relative to the baseline. In this plot, Density makes the highest contribution, adding +0.21 to the model's output, indicating its significant importance in the prediction.

Arom (wt%) and Mol Wt also make positive contributions of +0.07 and +0.06, respectively, further increasing the final output. On the other hand, P.CH (wt%) has a negative contribution of -0.01, slightly lowering the final predicted value. This step-by-step visualization helps in understanding the specific impact of each feature on the model's prediction.

4.8 Model Evaluation: MLs vs. Deep Learning

Based on results represented on **Table 4-4** it reveals that the deep Learning model ANN outperformed all other ML models, achieving the best predictive accuracy on both training and testing datasets.

Table 4-4: Comparison of metrics performance for ML and deep learning models for predicting SP

Model	RMSE (Train)	RMSE (Test)	MAE (Train)	MAE (Test)	R² (Train)	R² (Test)
ETR	0.74	2.47	0.51	1.37	0.97	0.79
XGBR	0.05	1.63	0.04	0.96	1	0.91
LightGBM	1.07	2.53	0.65	1.34	0.93	0.78
Nu-SVR	0.88	1.02	0.59	0.72	0.95	0.96
SVM	0.81	0.84	0.51	0.6	0.96	0.98
RF	0.83	2.32	0.54	1.31	0.96	0.82
ANN	0.18	0.23	0.13	0.18	0.96	0.97

With an R^2 value of 0.97 on the testing data and significantly low RMSE (0.23) and MAE (0.18), the Deep Learning Model demonstrated its capability to effectively capture complex, non-linear relationships in the dataset, making it the most reliable model for predicting smoke points. Among the ML models, SVM and Nu-SVR are considered strong alternatives, exhibiting balanced performance across all metrics. The SVM achieved the highest R^2 value (0.98) and a low RMSE (0.84) on testing data, indicating strong generalizability. Similarly, NuSVR maintained consistent results, with an R^2 of 0.96 and relatively low RMSE (1.02), reflecting its ability to generalize effectively while minimizing prediction errors. In contrast, tree-based models like XGBR and RF demonstrated limitations due to overfitting. While XGBR performed perfectly on the training data with the lowest RMSE (0.05) and MAE (0.04), its testing performance declined significantly, with an RMSE of 1.63 and an R^2 of 0.91. Random Forest showed a similar pattern, achieving high accuracy on training data (R^2 of 0.96) but struggling to generalize to unseen data, with an R^2 of 0.82 and an RMSE of 2.32 on testing data. The ETR and LightGBM also exhibited similar tendencies, with relatively high RMSE values (2.47 and 2.53, respectively) on testing data, indicating weaker predictive performance and challenges in generalizing across the dataset.

To sum up, the Deep Learning Model demonstrated the best performance and generalizability, SVM and Nu-SVR are viable alternatives for scenarios where computational simplicity and efficiency are critical. Tree-based models, despite their strengths in capturing relationships within training data, require careful tuning and extended dataset to avoid overfitting and improve their generalization capabilities.

CHAPTER 5

CONCLUSION

5.1 Conclusion

The research conducted indicates that adding oxygenated additives such as ethers, alcohols, and ketones to hydrocarbon-based fuels can effectively reduce soot emissions and improve combustion characteristics. According to experimental data, ethers—in particular, dibutyl ether (DBE)—achieved the greatest SP values when combined with TPRF80-20 because of the non-sooting fragments that their chemical structure produced. At greater mix levels, it has been demonstrated that straight-chain alcohols such as ethanol (EtOH) and n-pentanol (n-PeOH) greatly increase SP, with ethanol outperforming other alcohols in this regard. At lower concentrations, cyclopentanone (CPN) showed significant reactivity, lowering soot emissions, whereas at greater levels, cyclohexanone (CHN) burned cleaner. The complexity of additive behavior was shown by the different impacts that branching alcohols and ketones had on SP. As the TSI was measured for all oxygenated blends showing a reduction in soot tendency, focusing on 14% of oxygenated additives, ethanol showed the highest TSI reduction of 40% as seen in **Figure 5.1**.

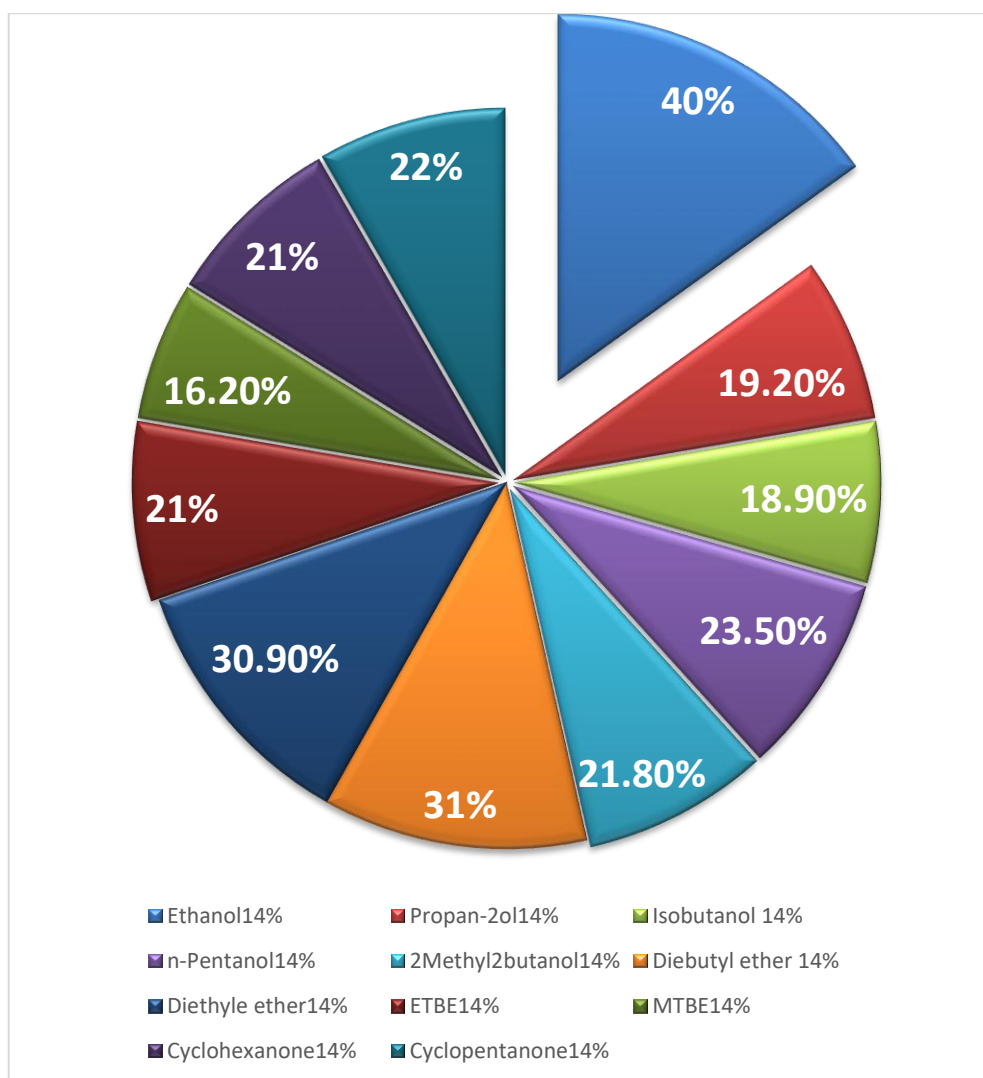


Figure 5.1: Soot tendency reduction for 14% oxygenated additives

As it came to predicting SP, the deep learning ANN model did better computationally than other ML models. With a Mean Absolute Error (MAE) of 0.18, R2 of 0.97, and Root Mean Squared Error (RMSE) of 0.23, the ANN demonstrated exceptional accuracy on both the testing and training datasets. Considering R2 values of 0.98 and 0.96, respectively, Support Vector Machine (SVM) and Nu-Support Vector Regression (Nu-SVR) outperformed the other ML models in terms of prediction.

However, overfitting was a problem for tree-based models like Random Forest and Light Gradient Boosting, which resulted in poorer generalization abilities.

results highlight the benefit of using deep learning methods to capture complicated and non-linear relationships in the dataset. The combined experimental and computational results demonstrate how advanced modelling techniques and properly designed oxygenated additions can improve combustion efficiency and lower soot emissions. The performance and environmental issues in contemporary combustion technologies are addressed by this comprehensive strategy, which advances cleaner fuel technology.

5.2 Future Work

- ❖ **Broadens Dataset Diversity:** To improve model generalizability, more fuel compositions, such as novel oxygenates and other blends based on hydrocarbons, should be included in future investigations.
- ❖ **Examine New Additive Classes:** To further optimize SP and lower particulate matter emissions, future studies may examine additional oxygenated molecules, such as esters, aldehydes and advanced biofuels.
- ❖ **Advanced Modelling Techniques:** To increase prediction accuracy and interpretability, use more complex algorithms, such as generative AI or hybrid ML-deep learning frameworks.

- ❖ Optimize Experimental Conditions: To gain a better understanding of the relationship between SP and soot generation, examine how different combustion parameters, such as temperature and pressure, affect both processes.
- ❖ Applications in the Actual World: Verify the models and experimental results in actual combustion systems in order to evaluate their scalability and effect on emissions and engine efficiency.

6 REFERENCES

- [1] CY8151-Engineering Chemistry-1908708516-unit_4 n.d.
- [2] Freitas RSM, Lima ÁPF, Chen C, Rochinha FA, Mira D, Jiang X. Towards predicting liquid fuel physicochemical properties using molecular dynamics guided machine learning models. *Fuel* 2022;329. <https://doi.org/10.1016/j.fuel.2022.125415>.
- [3] Han X, Jia M, Chang Y, Li Y, Wu S. Directed message passing neural network (D-MPNN) with graph edge attention (GEA) for property prediction of biofuel-relevant species. *Energy and AI* 2022;10. <https://doi.org/10.1016/j.egyai.2022.100201>.
- [4] Park Y, Moses-DeBusk M, Powell T, Szybist J, Xiang Z, Zhu J, et al. Fuel property impacts on gaseous and PM emissions from a multi-mode single-cylinder engine. *Fuel* 2023;331:125641. <https://doi.org/10.1016/J.FUEL.2022.125641>.
- [5] Ahmed Qasem MA, Al-Mutairi EM, Abdul Jameel AG. Smoke point prediction of oxygenated fuels using neural networks. *Fuel* 2023;332. <https://doi.org/10.1016/j.fuel.2022.126026>.
- [6] Zhang Y, Lou D, Tan P, Hu Z, Fang L. Effect of catalyzed diesel particulate filter and its catalyst loading on emission characteristics of a non-road diesel engine. *J Environ Sci (China)* 2023;126:794–805. <https://doi.org/10.1016/J.JES.2021.12.028>.
- [7] Zhang H, Li G, Long Y, Zhang Z, Wei W, Zhou M, et al. Numerical study on combustion and emission characteristics of a spark-ignition ammonia engine added with hydrogen-rich gas from exhaust-fuel reforming. *Fuel* 2023;332:125939. <https://doi.org/10.1016/J.FUEL.2022.125939>.
- [8] Ning W, Deng Q, Zhichun Y, - al, Mahdavi M, Hoar E, et al. You may also like Study of designable coefficient of thermal expansion in cellular structures Inverse modeling of inelastic properties of a two-phase microstructure Simultaneous two-step phase-shifting lateral shearing interferometry for aspherical surface based on an orthogonal shear displacer Evaluation of thermal properties of commercial fuels at high pressure and temperature through developed computer codes 2020. <https://doi.org/10.1088/2631-8695/aba8e2>.
- [9] 'Badra J, 'Pal P, 'Pei Y, 'Som S. Artificial Intelligence and Data Driven Optimization of Internal Combustion Engines. Amsterdam: Matthew Deans, Elsevier INC; 2022.
- [10] Rubio-Loyola J, Paul-Fils WRS. Applied Machine Learning in Industry 4.0: Case-Study Research in Predictive Models for Black Carbon Emissions. *Sensors* 2022;22. <https://doi.org/10.3390/s22103947>.
- [11] Samuel AL. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*. 1959;3.
- [12] Chang C-C, Chen S-H. A comparative analysis on artificial neural network-based two-stage clustering under a Creative Commons Attribution (CC-BY) 4.0 license A comparative analysis on artificial neural network-based two-stage clustering A comparative analysis on artificial neu 2015. <https://doi.org/10.1080/23311916.2014.995785>.
- [13] Adun H, Wole-Osho I, Okonkwo EC, Ruwa T, Agwa T, Onochie K, et al. Estimation of thermophysical property of hybrid nanofluids for solar Thermal applications: Implementation of novel Optimizable Gaussian Process regression (O-GPR) approach for Viscosity prediction. *Neural Comput Appl* 2022;34. <https://doi.org/10.1007/s00521-022-07038-2>.

- [14] I. Shyamkishore SM and RB. Alternative Transportation Fuels. Minutes of the Meeting - Pennsylvania Electric Association, Engineering Section 2022;1:1–16. <https://doi.org/10.30574/wjaets.2022.7.2.0123>.
- [15] Holechek JL, Geli HME, Sawalhah MN, Valdez R. A Global Assessment: Can Renewable Energy Replace Fossil Fuels by 2050? Sustainability (Switzerland) 2022;14:1–22. <https://doi.org/10.3390/su14084792>.
- [16] Pisupadti S V. Fuel Chemistry. Encyclopedia of Physical Science and Technology 2003;253–74. <https://doi.org/10.1016/B0-12-227410-5/00933-9>.
- [17] Xi J, Yang G, Cai J, Gu Z. A Review of Recent Research Results on Soot: The Formation of a Kind of Carbon-Based Material in Flames. Front Mater 2021;8:1–6. <https://doi.org/10.3389/fmats.2021.695485>.
- [18] Lindstedt RP, Michelsen HA, Mueller ME. Special issue and perspective on the chemistry and physics of carbonaceous particle formation. Combust Flame 2023;258:113042. <https://doi.org/10.1016/j.combustflame.2023.113042>.
- [19] Kent JH. A quantitative relationship between soot yield and smoke point measurements. Combust Flame 1986;63:349–58. [https://doi.org/10.1016/0010-2180\(86\)90004-0](https://doi.org/10.1016/0010-2180(86)90004-0).
- [20] Beji T, Zhang JP, Delichatsios M. Determination of Soot Formation Rate from Laminar Smoke Point Measurements. Combustion Science and Technology 2008;180:927–40. <https://doi.org/10.1080/00102200801894398>.
- [21] Saffaripour M, Zabeti P, Kholghy M, Thomson MJ. An experimental comparison of the sooting behavior of synthetic jet fuels. Energy and Fuels 2011;25:5584–93. https://doi.org/10.1021/EF201219V/ASSET/IMAGES/LARGE/EF-2011-01219V_0011.JPEG.
- [22] Pande SG, Hardy DR. An In-Depth Evaluation of Combustion Performance Predictors of Aviation Fuels Sooting Tendencies. Energy and Fuels 1995;9:448–57. https://doi.org/10.1021/EF00051A009/ASSET/EF00051A009.FP.PNG_V03.
- [23] Sunderland PB, Faeth GM. Soot formation in hydrocarbon/air laminar jet diffusion flames. Combust Flame 1996;105:132–46. [https://doi.org/10.1016/0010-2180\(95\)00182-4](https://doi.org/10.1016/0010-2180(95)00182-4).
- [24] Kent JH. A Quantitative Relationship Between Soot Yield and Smoke Point Measurements. vol. 63. 1986.
- [25] Ahmed Qasem MA, Al-Mutairi EM, Abdul Jameel AG. Smoke point prediction of oxygenated fuels using neural networks. Fuel 2023;332. <https://doi.org/10.1016/j.fuel.2022.126026>.
- [26] Drews A. Standard Test Method for Smoke Point of Kerosine and Aviation Turbine Fuel. Manual on Hydrocarbon Analysis, 6th Edition 2008:263–263–6. <https://doi.org/10.1520/mnl10868m>.
- [27] Olson DB, Pickens JC, Gill RJ. The effects of molecular structure on soot formation II. Diffusion flames. Combust Flame 1985;62:43–60. [https://doi.org/10.1016/0010-2180\(85\)90092-6](https://doi.org/10.1016/0010-2180(85)90092-6).
- [28] Gao Z, Zou X, Huang Z, Zhu L. Predicting sooting tendencies of oxygenated hydrocarbon fuels with machine learning algorithms. Fuel 2019;2420:438–46. <https://doi.org/10.1016/J.FUEL.2019.01.064>.
- [29] Calcote HF, Manos DM. Effect of molecular structure on incipient soot formation. Combust Flame 1983;49:289–304. [https://doi.org/10.1016/0010-2180\(83\)90172-4](https://doi.org/10.1016/0010-2180(83)90172-4).
- [30] Gill RJ, Olson DB. Estimation of Soot Thresholds for Fuel Mixtures. Combustion Science

- And Technology 1984;40:307–15. <https://doi.org/10.1080/00102208408923814>.
- [31] Barrientos EJ, Lapuerta M, Boehman AL. Group additivity in soot formation for the example of C-5 oxygenated hydrocarbon fuels. *Combust Flame* 2013;160:1484–98. <https://doi.org/10.1016/j.combustflame.2013.02.024>.
- [32] Das DD, St. John PC, McEnally CS, Kim S, Pfefferle LD. Measuring and predicting sooting tendencies of oxygenates, alkanes, alkenes, cycloalkanes, and aromatics on a unified scale. *Combust Flame* 2018;190:349–64. <https://doi.org/10.1016/J.COMBUSTFLAME.2017.12.005>.
- [33] McEnally CS, Pfefferle LD. Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways. *Combust Flame* 2007;148:210–22. <https://doi.org/10.1016/J.COMBUSTFLAME.2006.11.003>.
- [34] Alboqami FD, Pasha AA, Alam MI, Abdulraheem A, Abdul Jameel AG. Prediction of Yield Sooting Index Utilizing Artificial Neural Networks and Adaptive-Network-Based Fuzzy Inference Systems. *Arab J Sci Eng* 2023;48:8901–9. <https://doi.org/10.1007/S13369-022-07561-3/METRICS>.
- [35] McEnally CS, Pfefferle LD. Sooting tendencies of nonvolatile aromatic hydrocarbons. *Proceedings of the Combustion Institute* 2009;32:673–9. <https://doi.org/10.1016/J.PROCI.2008.06.197>.
- [36] McEnally CS, Pfefferle LD. Sooting tendencies of oxygenated hydrocarbons in laboratory-scale flames. *Environ Sci Technol* 2011;45:2498–503. https://doi.org/10.1021/ES103733Q/SUPPL_FILE/ES103733Q_SI_001.PDF.
- [37] Das DD, McEnally CS, Pfefferle LD. Sooting tendencies of unsaturated esters in nonpremixed flames. *Combust Flame* 2015;162:1489–97. <https://doi.org/10.1016/J.COMBUSTFLAME.2014.11.012>.
- [38] Crossley SP, Alvarez WE, Resasco DE. Novel Micropyrolysis Index (MPI) to Estimate the Sooting Tendency of Fuels. *Energy and Fuels* 2008;22:2455–64. <https://doi.org/10.1021/EF800058Y>.
- [39] Lemaire R, Lapalme D, Seers P. Analysis of the sooting propensity of C-4 and C-5 oxygenates: Comparison of sooting indexes issued from laser-based experiments and group additivity approaches. *Combust Flame* 2015;162:3140–55. <https://doi.org/10.1016/j.combustflame.2015.03.018>.
- [40] Lemaire R, Le Corre G, Nakouri M. Predicting the propensity to soot of hydrocarbons and oxygenated molecules by means of structural group contribution factors derived from the processing of unified sooting indexes. *Fuel* 2021;302. <https://doi.org/10.1016/j.fuel.2021.121104>.
- [41] Yang Y, Boehman AL, Santoro RJ. A study of jet fuel sooting tendency using the threshold sooting index (TSI) model. *Combust Flame* 2007;149:191–205. <https://doi.org/10.1016/j.combustflame.2006.11.007>.
- [42] Ahmed Qasem MA, van Oudenhoven VCO, Pasha AA, Pillai SN, Reddy VM, Ahmed U, et al. A machine learning model for predicting threshold sooting index (TSI) of fuels containing alcohols and ethers. *Fuel* 2022;322:123941. <https://doi.org/10.1016/J.FUEL.2022.123941>.
- [43] Ladommatos N, Rubenstein P, Bennett P. Some effects of molecular structure of single hydrocarbons on sooting tendency. *Fuel* 1996;75:114–24. [https://doi.org/10.1016/0016-2361\(94\)00251-7](https://doi.org/10.1016/0016-2361(94)00251-7).
- [44] Allan KM, Kaminski JR, Bertrand JC, Head J, Sunderland PB. Laminar smoke points of wax candles. *Combustion Science and Technology* 2009;181:800–11.

- <https://doi.org/10.1080/00102200902935512>.
- [45] Pepiot-Desjardins P, Pitsch H, Malhotra R, Kirby SR, Boehman AL. Structural group analysis for soot reduction tendency of oxygenated fuels. *Combust Flame* 2008;154:191–205. <https://doi.org/10.1016/J.COMBUSTFLAME.2008.03.017>.
 - [46] Haas FM, Qin A, Dryer FL. “Virtual” smoke point determination of alternative aviation kerosenes by threshold sooting index (TSI) methods. 50th AIAA/ASME/SAE/ASEE Joint Propulsion Conference 2014 2014:1–9. <https://doi.org/10.2514/6.2014-3468>.
 - [47] Eliaz J, Osipowicz T, Abramek KF, Matuszak Z, Mozga Ł. Fuel Pretreatment Systems in Modern CI Engines. *Catalysts* 2020, Vol 10, Page 696 2020;10:696. <https://doi.org/10.3390/CATAL10060696>.
 - [48] Jo S, Park S, Kim HJ, Lee JT. Combustion improvement and emission reduction through control of ethanol ratio and intake air temperature in reactivity controlled compression ignition combustion engine. *Appl Energy* 2019;250:1418–31. <https://doi.org/10.1016/J.APENERGY.2019.05.012>.
 - [49] Zhang F, Wang Z, Tian J, Li L, Yu K, He K. Effect of EGR and Fuel Injection Strategies on the Heavy-Duty Diesel Engine Emission Performance under Transient Operation. *Energies* 2020, Vol 13, Page 566 2020;13:566. <https://doi.org/10.3390/EN13030566>.
 - [50] Schripp T, Anderson B, Crosbie EC, Moore RH, Herrmann F, Oßwald P, et al. Impact of Alternative Jet Fuels on Engine Exhaust Composition during the 2015 ECLIF Ground-Based Measurements Campaign. *Environ Sci Technol* 2018;52:4969–78. https://doi.org/10.1021/ACS.EST.7B06244/SUPPL_FILE/ES7B06244_SI_001.PDF.
 - [51] Rodríguez-Fernández J, Hernández JJ, Sánchez-Valdepeñas J. Effect of oxygenated and paraffinic alternative diesel fuels on soot reactivity and implications on DPF regeneration. *Fuel* 2016;185:460–7. <https://doi.org/10.1016/J.FUEL.2016.08.016>.
 - [52] Han Y, Chen Y, Feng Y, Shang Y, Li J, Li Q, et al. Fuel aromaticity promotes low-temperature nucleation processes of elemental carbon from biomass and coal combustion. *Environ Sci Technol* 2021;55:2532–40. https://doi.org/10.1021/ACS.EST.0C06694/SUPPL_FILE/ES0C06694_SI_001.PDF.
 - [53] Park W, Park S, Reitz RD, Kurtz E. The effect of oxygenated fuel properties on diesel spray combustion and soot formation. *Combust Flame* 2017;180:276–83. <https://doi.org/10.1016/J.COMBUSTFLAME.2016.02.026>.
 - [54] Moon S, Tsujimura T, Oguma M, Chen Z, Huang Z, Saitou T. Mixture condition, combustion and sooting characteristics of ethanol–diesel blends in diffusion flames under various injection and ambient conditions. *Fuel* 2013;113:128–39. <https://doi.org/10.1016/J.FUEL.2013.05.060>.
 - [55] An H, Yang WM, Li J, Zhou DZ. Modeling study of oxygenated fuels on diesel combustion: Effects of oxygen concentration, cetane number and C/H ratio. *Energy Convers Manag* 2015;90:261–71. <https://doi.org/10.1016/J.ENCONMAN.2014.11.031>.
 - [56] Marshall AG, Rodgers RP. Petroleomics: Chemistry of the underworld. *Proc Natl Acad Sci U S A* 2008;105:18090–5. <https://doi.org/10.1073/PNAS.0805069105/ASSET/09C576C9-AEC2-4060-AD43-72E5FC6E7982/ASSETS/GRAPHIC/ZPQ9990850000006.JPEG>.
 - [57] Abdul Jameel AG, Alquaity ABS, Campuzano F, Emwas A-H, Saxena S, Sarathy SM, et al. Surrogate formulation and molecular characterization of sulfur species in vacuum residues using APPI and ESI FT-ICR mass spectrometry. *Fuel* 2021;293. <https://doi.org/10.1016/j.fuel.2021.120471>.
 - [58] Abdul Jameel AG, Naser N, Issayev G, Touitou J, Ghosh MK, Emwas AH, et al. A

- minimalist functional group (MFG) approach for surrogate fuel formulation. *Combust Flame* 2018;192:250–71. <https://doi.org/10.1016/J.COMBUSTFLAME.2018.01.036>.
- [59] Abdul Jameel AG, Elbaz AM, Emwas AH, Roberts WL, Sarathy SM. Calculation of Average Molecular Parameters, Functional Groups, and a Surrogate Molecule for Heavy Fuel Oils Using ¹H and ¹³C Nuclear Magnetic Resonance Spectroscopy. *Energy and Fuels* 2016;30:3894–905. <https://doi.org/10.1021/acs.energyfuels.6b00303>.
- [60] Thuy T Le, Allen DT. Property estimation using structural characterizations: Heat capacities of coal liquids. *Fuel* 1985;64:1754–9. [https://doi.org/10.1016/0016-2361\(85\)90404-1](https://doi.org/10.1016/0016-2361(85)90404-1).
- [61] Benson SW, Golden DM, Haugen GR, Shaw R, Cruickshank FR, Rodgers AS, et al. Additivity rules for the estimation of thermochemical properties. *Chem Rev* 1969;69:279–324. https://doi.org/10.1021/CR60259A002/ASSET/CR60259A002.FP.PNG_V03.
- [62] Saldana DA, Starck L, Mougin P, Rousseau B, Pidol L, Jeuland N, et al. Flash point and cetane number predictions for fuel compounds using quantitative structure property relationship (QSPR) methods. *Energy and Fuels* 2011;25:3900–8. <https://doi.org/10.1021/ef200795j>.
- [63] Álvarez A, Lapuerta M, Agudelo JR. Prediction of Flash-Point Temperature of Alcohol/Biodiesel/Diesel Fuel Blends. *Ind Eng Chem Res* 2019;58:6860–9. <https://doi.org/10.1021/acs.iecr.9b00843>.
- [64] Saldana DA, Starck L, Mougin P, Rousseau B, Creton B. Prediction of flash points for fuel mixtures using machine learning and a novel equation. *Energy and Fuels* 2013;27:3811–20. <https://doi.org/10.1021/ef4005362>.
- [65] Aljaman B, Ahmed U, Zahid U, Reddy VM, Sarathy SM, Abdul Jameel AG. A comprehensive neural network model for predicting flash point of oxygenated fuels using a functional group approach. *Fuel* 2022;317:123428. <https://doi.org/10.1016/j.fuel.2022.123428>.
- [66] Yuan H, Yang B, Yang J. Predicting properties of biodiesel fuels using mixture topological index. *JAOCS, Journal of the American Oil Chemists' Society* 2009;86:375–82. <https://doi.org/10.1007/S11746-009-1354-Y/METRICS>.
- [67] Stratiev D, Marinov I, Dinkov R, Shishkova I, Velkov I, Sharafutdinov I, et al. Opportunity to improve diesel-fuel cetane-number prediction from easily available physical properties and application of the least-squares method and artificial neural networks. *Energy and Fuels* 2015;29:1520–33. <https://doi.org/10.1021/ef502638c>.
- [68] O'Connor CT, Forrester RD, Scurrrell MS. Cetane number determination of synthetic diesel fuels. *Fuel* 1992;71:1323–7. [https://doi.org/10.1016/0016-2361\(92\)90061-R](https://doi.org/10.1016/0016-2361(92)90061-R).
- [69] Creton B, Dartiguelongue C, De Bruin T, Toulhoat H. Prediction of the cetane number of diesel compounds using the quantitative structure property relationship. *Energy and Fuels* 2010;24:5396–403. <https://doi.org/10.1021/ef1008456>.
- [70] Won SH, Dooley S, Veloo PS, Wang H, Oehlschlaeger MA, Dryer FL, et al. The combustion properties of 2,6,10-trimethyl dodecane and a chemical functional group analysis. *Combust Flame* 2014;161:826–34. <https://doi.org/10.1016/J.COMBUSTFLAME.2013.08.010>.
- [71] Heck SM, Pritchard HO, Griffiths JF. Cetane number vs. structure in paraffin hydrocarbons. *Journal of the Chemical Society - Faraday Transactions* 1998;94:1725–7. <https://doi.org/10.1039/a800861b>.
- [72] Dooley S, Won SH, Heyne J, Farouk TI, Ju Y, Dryer FL, et al. The experimental

- evaluation of a methodology for surrogate fuel formulation to emulate gas phase combustion kinetic phenomena. *Combust Flame* 2012;159:1444–66.
<https://doi.org/10.1016/J.COMBUSTFLAME.2011.11.002>.
- [73] Lapuerta M, Hernández JJ, Sarathy SM. Effects of methyl substitution on the auto-ignition of C16 alkanes. *Combust Flame* 2016;164:259–69.
<https://doi.org/10.1016/J.COMBUSTFLAME.2015.11.024>.
- [74] Hernández-Trujillo J, Bader RFW. Properties of atoms in molecules: Atoms forming molecules. *Journal of Physical Chemistry A* 2000;104:1779–94.
<https://doi.org/10.1021/JP994096Z/ASSET/IMAGES/MEDIUM/JP994096ZE00010.GIF>.
- [75] Sumathi R, Carstensen HH, Green WH. Reaction rate prediction via group additivity part 1: H abstraction from alkanes by H and CH₃. *Journal of Physical Chemistry A* 2001;105:6910–25.
https://doi.org/10.1021/JP010697Q/SUPPL_FILE/JP010697Q_S.PDF.
- [76] Sumathi R, Carstensen HH, Green WH. Reaction rate prediction via group additivity, part 2: H-abstraction from alkenes, alkynes, alcohols, aldehydes, and acids by H atoms. *Journal of Physical Chemistry A* 2001;105:8969–84.
https://doi.org/10.1021/JP011827Y/SUPPL_FILE/JP011827Y_S.PDF.
- [77] Sumathi R, Carstensen HH, Green WH. Reaction rate predictions via group additivity. Part 3: Effect of substituents with CH₂ as the mediator. *Journal of Physical Chemistry A* 2002;106:5474–89.
https://doi.org/10.1021/JP013957C/SUPPL_FILE/JP013957C_S.PDF.
- [78] Pepiot-Desjardins P, Pitsch H, Malhotra R, Kirby SR, Boehman AL. Structural group analysis for soot reduction tendency of oxygenated fuels. *Combust Flame* 2008;154:191–205. <https://doi.org/10.1016/J.COMBUSTFLAME.2008.03.017>.
- [79] Yang Y, Boehman AL, Santoro RJ. A study of jet fuel sooting tendency using the threshold sooting index (TSI) model. *Combust Flame* 2007;149:191–205.
<https://doi.org/10.1016/j.combustflame.2006.11.007>.
- [80] Mehl M, Pitz WJ, Sarathy SM, Westbrook CK. Modeling the combustion of high molecular weight fuels by a functional group approach. *Int J Chem Kinet* 2012;44:257–76. <https://doi.org/10.1002/KIN.20715>.
- [81] Yan S, Eddings EG, Palotas AB, Pugmire RJ, Sarofim AF. Prediction of sooting tendency for hydrocarbon liquids in diffusion flames. *Energy and Fuels* 2005;19:2408–15.
<https://doi.org/10.1021/EF050107D/ASSET/IMAGES/LARGE/EF050107DF00006.JPEG>.
- [82] Wang W, Li Z, Oehlschlaeger MA, Healy D, Curran HJ, Sarathy SM, et al. An experimental and modeling study of the autoignition of 3-methylheptane. *Proceedings of the Combustion Institute* 2013;34:335–43.
<https://doi.org/10.1016/J.PROCI.2012.06.001>.
- [83] Sarathy SM, Westbrook CK, Mehl M, Pitz WJ, Togbe C, Dagaut P, et al. Comprehensive chemical kinetic modeling of the oxidation of 2-methylalkanes from C₇ to C₂₀. *Combust Flame* 2011;158:2338–57.
<https://doi.org/10.1016/J.COMBUSTFLAME.2011.05.007>.
- [84] Selim H, Mohamed SY, Lucassen A, Hansen N, Sarathy SM. Effect of the methyl substitution on the combustion of two methylheptane isomers: Flame chemistry using vacuum-ultraviolet (VUV) photoionization mass spectrometry. *Energy and Fuels* 2015;29:2696–708.
https://doi.org/10.1021/EF502797A/SUPPL_FILE/EF502797A_SI_002.ZIP.

- [85] Sarathy SM, Yeung C, Westbrook CK, Pitz WJ, Mehl M, Thomson MJ. An experimental and kinetic modeling study of n-octane and 2-methylheptane in an opposed-flow diffusion flame. *Combust Flame* 2011;158:1277–87. <https://doi.org/10.1016/J.COMBUSTFLAME.2010.11.008>.
- [86] Karsenty F, Sarathy SM, Togbé C, Westbrook CK, Dayma G, Dagaut P, et al. Experimental and kinetic modeling study of 3-methylheptane in a jet-stirred reactor. *Energy and Fuels* 2012;26:4680–9. https://doi.org/10.1021/EF300852W/SUPPL_FILE/EF300852W_SI_001.ZIP.
- [87] Abdul Jameel AG, Naser N, Emwas AH, Dooley S, Sarathy SM. Predicting Fuel Ignition Quality Using 1H NMR Spectroscopy and Multiple Linear Regression. *Energy and Fuels* 2016;30:9819–35. <https://doi.org/10.1021/acs.energyfuels.6b01690>.
- [88] Kang D, Lilik G, Dillstrom V, Agudelo J, Lapuerta M, Al-Qurashi K, et al. Impact of branched structures on cycloalkane ignition in a motored engine: Detailed product and conformational analyses. *Combust Flame* 2015;162:877–92. <https://doi.org/10.1016/J.COMBUSTFLAME.2014.09.009>.
- [89] Kang D, Kirby S, Agudelo J, Lapuerta M, Al-Qurashi K, Boehman AL. Combined impact of branching and unsaturation on the autoignition of binary blends in a motored engine. *Energy and Fuels* 2014;28:7203–15. https://doi.org/10.1021/EF501629P/ASSET/IMAGES/LARGE/EF-2014-01629P_0017.JPEG.
- [90] Abdul Jameel AG, Oudenhoven V Van, Emwas AH, Sarathy SM. Predicting Octane Number Using Nuclear Magnetic Resonance Spectroscopy and Artificial Neural Networks. *Energy and Fuels* 2018;32:6309–29. <https://doi.org/10.1021/acs.energyfuels.8b00556>.
- [91] Curran HJ, Gaffuri P, Pitz WJ, Westbrook CK. A comprehensive modeling study of n-heptane oxidation. *Combust Flame* 1998;114:149–77. [https://doi.org/10.1016/S0010-2180\(97\)00282-4](https://doi.org/10.1016/S0010-2180(97)00282-4).
- [92] Curran HJ, Gaffuri P, Pitz WJ, Westbrook CK. A comprehensive modeling study of iso-octane oxidation. *Combust Flame* 2002;129:253–80. [https://doi.org/10.1016/S0010-2180\(01\)00373-X](https://doi.org/10.1016/S0010-2180(01)00373-X).
- [93] Gani A, Jameel A. A functional group approach for predicting fuel properties. 2019.
- [94] Kumar M, Karmakar S, Nimesh V. Statistical investigation of combustion and emission characteristics of biofuels according to their physical properties: A way to explore suitable alternative fuels. *Fuel* 2024;358:130242. <https://doi.org/10.1016/J.FUEL.2023.130242>.
- [95] Pischinger S, Rajamani VK, Jekhouri Y. Impact of fuel properties on the performance of a direct injection diesel engine under part homogeneous operating conditions. *SAE 2011 World Congress and Exhibition*, 2011. <https://doi.org/10.4271/2011-01-1358>.
- [96] Xu L, Wang Y, Liu D. Effects of oxygenated biofuel additives on soot formation: A comprehensive review of laboratory-scale studies. *Fuel* 2022;313:122635. <https://doi.org/10.1016/j.fuel.2021.122635>.
- [97] Kalghatgi G, Head R, Chang J, Viollet Y, Babiker H, Amer A. An Alternative Method Based on Toluene/n-Heptane Surrogate Fuels for Rating the Anti-Knock Quality of Practical Gasolines. *SAE Int J Fuels Lubr* 2014;7:663–72. <https://doi.org/10.4271/2014-01-2609>.
- [98] Kalghatgi G, Babiker H, Badra J. A Simple Method to Predict Knock Using Toluene, N-Heptane and Iso-Octane Blends (TPRF) as Gasoline Surrogates. *SAE Int J Engines*

- 2015;8:505–19. <https://doi.org/10.4271/2015-01-0757>.
- [99] Ahmed A, Goteng G, Shankar VSB, Al-Qurashi K, Roberts WL, Sarathy SM. A computational methodology for formulating gasoline surrogate fuels with accurate physical and chemical kinetic properties. *Fuel* 2015;143:290–300. <https://doi.org/10.1016/j.fuel.2014.11.022>.
- [100] Kim D, Martz J, Violi A. A surrogate for emulating the physical and chemical properties of conventional jet fuel. *Combust Flame* 2014;161:1489–98. <https://doi.org/10.1016/j.combustflame.2013.12.015>.
- [101] Bhavani Shankar VS, Sajid M, Al-Qurashi K, Atef N, Alkhesho I, Ahmed A, et al. Primary Reference Fuels (PRFs) as Surrogates for Low Sensitivity Gasoline Fuels. *SAE Technical Papers* 2016. <https://doi.org/10.4271/2016-01-0748>.
- [102] Waqas MU, Mohammed A, Masurier JB, Johansson B. Blending Octane Number of 1-Butanol and Iso-Octane with Low Octane Fuels in HCCI Combustion Mode. *SAE Technical Papers* 2018;2018-Sept:1–11. <https://doi.org/10.4271/2018-01-1681>.
- [103] Javed T, Lee C, AlAbbad M, Djebbi K, Beshir M, Badra J, et al. Ignition studies of n-heptane/iso-octane/toluene blends. *Combust Flame* 2016;171:223–33. <https://doi.org/10.1016/j.combustflame.2016.06.008>.
- [104] McEnally CS, Pfefferle LD. The effects of dimethyl ether and ethanol on benzene and soot formation in ethylene nonpremixed flames. *Proceedings of the Combustion Institute* 2007;31:603–10. <https://doi.org/10.1016/j.proci.2006.07.005>.
- [105] Liu F, He X, Ma X, Zhang Q, Thomson MJ, Guo H, et al. An experimental and numerical study of the effects of dimethyl ether addition to fuel on polycyclic aromatic hydrocarbon and soot formation in laminar coflow ethylene/air diffusion flames. *Combust Flame* 2011;158:547–63. <https://doi.org/10.1016/J.COMBUSTFLAME.2010.10.005>.
- [106] Wullenkord J, Graf I, Baroncelli M, Felsmann D, Cai L, Pitsch H, et al. Laminar premixed and non-premixed flame investigation on the influence of dimethyl ether addition on n-heptane combustion. *Combust Flame* 2020;212:323–36. <https://doi.org/10.1016/J.COMBUSTFLAME.2019.11.012>.
- [107] Chen M, Liu D. Morphology and nanostructure transitions of soot with various dimethyl ether additions in nonpremixed ethylene flames at different scales. *Energy and Fuels* 2020;34:16705–19. <https://doi.org/10.1021/ACS.ENERGYFUELS.0C03305>/ASSET/IMAGES/LARGE/EF0C03305_0018.JPEG.
- [108] Yan F, Xu L, Wang Y, Park S, Sarathy SM, Chung SH. On the opposing effects of methanol and ethanol addition on PAH and soot formation in ethylene counterflow diffusion flames. *Combust Flame* 2019;202:228–42. <https://doi.org/10.1016/J.COMBUSTFLAME.2019.01.020>.
- [109] Zhang Y, Li Y, Liu P, Zhan R, Huang Z, Lin H. Investigation on the chemical effects of dimethyl ether and ethanol additions on PAH formation in laminar premixed ethylene flames. *Fuel* 2019;256:115809. <https://doi.org/10.1016/J.FUEL.2019.115809>.
- [110] Kang Y, Sun Y, Lu X, Gou X, Sun S, Yan J, et al. Soot formation characteristics of ethylene premixed burner-stabilized stagnation flame with dimethyl ether addition. *Energy* 2018;150:709–21. <https://doi.org/10.1016/J.ENERGY.2018.03.025>.
- [111] Jin H, Pieper J, Hemken C, Bräuer E, Ruwe L, Kohse-Höinghaus K. Chemical interaction of dual-fuel mixtures in low-temperature oxidation, comparing n-pentane/dimethyl ether and n-pentane/ethanol. *Combust Flame* 2018;193:36–53.

- <https://doi.org/10.1016/J.COMBUSTFLAME.2018.03.003>.
- [112] Sirignano M, Ciajolo A, D'Anna A, Russo C. Chemical Features of Particles Generated in an Ethylene/Ethanol Premixed Flame. *Energy and Fuels* 2017;31:2370–7. <https://doi.org/10.1021/ACS.ENERGYFUELS.6B02372>/ASSET/IMAGES/LARGE/EF-2016-02372J_0008.JPEG.
 - [113] Khosousi A, Liu F, Dworkin SB, Eaves NA, Thomson MJ, He X, et al. Experimental and numerical study of soot formation in laminar coflow diffusion flames of gasoline/ethanol blends. *Combust Flame* 2015;162:3925–33. <https://doi.org/10.1016/J.COMBUSTFLAME.2015.07.029>.
 - [114] Martin JA, Ratcliff MA, Rahimi MJ, Burton JL, Sindler P, Hays CK, et al. ϕ -Sensitivity of Gasoline/Oxygenate Blends in an Advanced Compression Ignition Engine. *Energy and Fuels* 2023;37:12243–58. <https://doi.org/10.1021/ACS.ENERGYFUELS.3C01537>/ASSET/IMAGES/LARGE/EF3C01537_0019.JPEG.
 - [115] Vishwanath RB, Carniglia PA, Weber JK, Gülder ÖL. Effects of n-pentanol blending on soot formation in swirl-stabilized turbulent spray flames of Jet A-1 in a laboratory gas turbine combustor. *Fuel* 2024;357:129971. <https://doi.org/10.1016/J.FUEL.2023.129971>.
 - [116] Wei L, Cheung CS, Huang Z. Effect of n-pentanol addition on the combustion, performance and emission characteristics of a direct-injection diesel engine. *Energy* 2014;70:172–80. <https://doi.org/10.1016/j.energy.2014.03.106>.
 - [117] Patil V, Singh P, Sonage S, Kumbhakarna N, Kumar S. Applicability of ketone-gasoline blended fuels for spark ignition engine through energy-exergy analyses. *Fuel* 2023;339:127416. <https://doi.org/10.1016/j.fuel.2023.127416>.
 - [118] Singh AP, Sharma N, Kumar V, Satsangi DP, Agarwal AK. Fuel Injection Strategy for Utilization of Mineral Diesel-Methanol Blend in a Common Rail Direct Injection Engine. *Journal of Energy Resources Technology, Transactions of the ASME* 2020;142:1–13. <https://doi.org/10.1115/1.4046225>.
 - [119] Karunamurthy K, Janvekar AA, Palaniappan PL, Adhitya V, Lokeswar TTK, Harish J. Prediction of IC engine performance and emission parameters using machine learning: A review. *J Therm Anal Calorim* 2023;148:3155–77. <https://doi.org/10.1007/S10973-022-11896-2>/METRICS.
 - [120] Veza I, Afzal A, Mujtaba MA, Tuan Hoang A, Balasubramanian D, Sekar M, et al. Review of artificial neural networks for gasoline, diesel and homogeneous charge compression ignition engine. *Alexandria Engineering Journal* 2022;61:8363–91. <https://doi.org/10.1016/J.AEJ.2022.01.072>.
 - [121] Kuzhagaliyeva N, Horváth S, Williams J, Nicolle A, Sarathy SM. Artificial intelligence-driven design of fuel mixtures. *Communications Chemistry* 2022 5:1 2022;5:1–10. <https://doi.org/10.1038/s42004-022-00722-3>.
 - [122] Alrefaai MM, Guerrero Peña GDJ, Raj A, Stephen S, Anjana T, Dindi A. Impact of dicyclopentadiene addition to diesel on cetane number, sooting propensity, and soot characteristics. *Fuel* 2018;216:110–20. <https://doi.org/10.1016/j.fuel.2017.11.145>.
 - [123] Abboud J, Schobing J, Legros G, Bonnetty J, Tschamber V, Brillard A, et al. Impacts of oxygenated compounds concentration on sooting propensities and soot oxidative reactivity: Application to Diesel and Biodiesel surrogates. *Fuel* 2017;193:241–53. <https://doi.org/10.1016/j.fuel.2016.12.034>.
 - [124] Abdul Jameel AG. Predicting sooting propensity of oxygenated fuels using artificial

- neural networks. *Processes* 2021;9:1070. <https://doi.org/10.3390/pr9061070>.
- [125] Ahmed Qasem MA, van Oudenhoven VCO, Pasha AA, Pillai SN, Reddy VM, Ahmed U, et al. A machine learning model for predicting threshold sooting index (TSI) of fuels containing alcohols and ethers. *Fuel* 2022;322:123941. <https://doi.org/10.1016/J.FUEL.2022.123941>.
- [126] Abdul Jameel AG. Identification and Quantification of Hydrocarbon Functional Groups in Gasoline Using ¹H-NMR Spectroscopy for Property Prediction. *Molecules* 2021;26:1–10. <https://doi.org/10.3390/molecules26226989>.
- [127] Keshavaraj R, Tock RW, Narayan RS, Bartsch RA, Ethridge D, Zhu R. Fluid property prediction of siloxanes with the aid of artificial neural nets. *Polymer - Plastics Technology and Engineering* 1996;35:971–82. <https://doi.org/10.1080/03602559608000611>.
- [128] Hall LH, Story CT. Boiling point and critical temperature of a heterogeneous data set: QSAR with atom type electrotopological state indices using artificial neural networks. *J Chem Inf Comput Sci* 1996;36:1004–14. <https://doi.org/10.1021/ci960375x>.
- [129] McClelland HE, Jurs PC. Quantitative Structure-Property Relationships for the Prediction of Vapor Pressures of Organic Compounds from Molecular Structures. *J Chem Inf Comput Sci* 2000;40:967–75. <https://doi.org/10.1021/ci990137c>.
- [130] Pan Y, Jiang J, Wang R, Cao H, Zhao J. Quantitive structure - Property relationship studies for predicting flash points of organic compounds using support vector machines. *QSAR Comb Sci* 2008;27:1013–9. <https://doi.org/10.1002/qsar.200810009>.
- [131] Basu S, Henshaw PF. Correlations of relative adsorbability with physical and structural properties of VOCs - Application of an artificial neural network. 2000 Annual Conference Abstracts - Canadian Society for Civil Engineering 2000:145.
- [132] Yalamanchi KK, Nicolle A, Sarathy SM. Artificial intelligence-enabled fuel design. *Artificial Intelligence and Data Driven Optimization of Internal Combustion Engines*, Elsevier; 2022, p. 47–67. <https://doi.org/10.1016/B978-0-323-88457-0.00011-4>.
- [133] 'Badra J, 'Pal P, 'Pei Y, 'Som S. Artificial Intelligence and Data Driven Optimization of Internal Combustion Engines. Amsterdam: Matthew Deans, Elsevier INC; 2022.
- [134] Abdul Jameel AG. Identification and Quantification of Hydrocarbon Functional Groups in Gasoline Using ¹H-NMR Spectroscopy for Property Prediction. *Molecules* 2021;26:1–10. <https://doi.org/10.3390/molecules26226989>.
- [135] Heck SM, Pritchard HO, Griffiths JF. Cetane number vs. structure in paraffin hydrocarbons. *Journal of the Chemical Society - Faraday Transactions* 1998;94:1725–7. <https://doi.org/10.1039/a800861b>.
- [136] Creton B, Dartiguelongue C, De Bruin T, Toulhoat H. Prediction of the cetane number of diesel compounds using the quantitative structure property relationship. *Energy and Fuels* 2010;24:5396–403. <https://doi.org/10.1021/ef1008456>.
- [137] Abdul Jameel AG, Naser N, Emwas AH, Dooley S, Sarathy SM. Predicting Fuel Ignition Quality Using ¹H NMR Spectroscopy and Multiple Linear Regression. *Energy and Fuels* 2016;30:9819–35. <https://doi.org/10.1021/acs.energyfuels.6b01690>.
- [138] Lapidus AL, Smolenskii EA, Bavykin VM, Myshenkova TN, Kondrat'Ev LT. Models for the calculation and prediction of the octane and cetane numbers of individual hydrocarbons. *Petroleum Chemistry* 2008;48:277–86. <https://doi.org/10.1134/S0965544108040051>.
- [139] Gülder OL, Glavinčevski B. Comments on "Prediction of Cetane Number by Group Additivity and Carbon-13 Nuclear Magnetic Resonance." *Ind Eng Chem Res*

- 1988;27:2192–4. <https://doi.org/10.1021/ie00083a043>.
- [140] Saldana DA, Starck L, Mougin P, Rousseau B, Pidol L, Jeuland N, et al. Flash point and cetane number predictions for fuel compounds using quantitative structure property relationship (QSPR) methods. *Energy and Fuels* 2011;25:3900–8. <https://doi.org/10.1021/ef200795j>.
- [141] Nadai D V., Simões JB, Gatts CEN, Miranda PCML. Inference of the biodiesel cetane number by multivariate techniques. *Fuel* 2013;105:325–30. <https://doi.org/10.1016/J.FUEL.2012.06.018>.
- [142] Kubic WL, Jenkins RW, Moore CM, Semelsberger TA, Sutton AD. Artificial Neural Network Based Group Contribution Method for Estimating Cetane and Octane Numbers of Hydrocarbons and Oxygenated Organic Compounds. *Ind Eng Chem Res* 2017;56:12236–45. <https://doi.org/10.1021/acs.iecr.7b02753>.
- [143] Schweidtmann AM, Rittig JG, König A, Grohe M, Mitsos A, Dahmen M. Graph Neural Networks for Prediction of Fuel Ignition Quality. *Energy and Fuels* 2020;34:11395–407. <https://doi.org/10.1021/acs.energyfuels.0c01533>.
- [144] Souza CR, Silva AH, Nagata N, Ribas JLT, Simonelli F, Barison A. Cetane number assessment in diesel fuel by ¹H or hydrogen nuclear magnetic resonance-based multivariate calibration. *Energy and Fuels* 2014;28:4958–62. <https://doi.org/10.1021/ef501000a>.
- [145] Kapur GS, Ecker A, Meusinger R. Establishing quantitative structure-property relationships (QSPR) of diesel samples by proton-NMR & multiple linear regression (MLR) analysis. *Energy and Fuels* 2001;15:943–8. <https://doi.org/10.1021/ef010021u>.
- [146] Alves JCL, Henriques CB, Poppi RJ. Determination of diesel quality parameters using support vector regression and near infrared spectroscopy for an in-line blending optimizer system. *Fuel* 2012;97:710–7. <https://doi.org/10.1016/j.fuel.2012.03.016>.
- [147] Alizadeh G, Frounchi J, Baradaran Nia M, Zarifi MH, Asgarifar S. An FPGA implementation of an Artificial Neural Network for prediction of cetane number. *Proceedings of the International Conference on Computer and Communication Engineering 2008, ICCCE08: Global Links for Human Development 2008*:605–8. <https://doi.org/10.1109/ICCCE.2008.4580675>.
- [148] Kessler T, Sacia ER, Bell AT, Mack JH. Artificial neural network based predictions of cetane number for furanic biofuel additives. *Fuel* 2017;206:171–9. <https://doi.org/10.1016/j.fuel.2017.06.015>.
- [149] O'Connor CT, Forrester RD, Scurrrell MS. Cetane number determination of synthetic diesel fuels. *Fuel* 1992;71:1323–7. [https://doi.org/10.1016/0016-2361\(92\)90061-R](https://doi.org/10.1016/0016-2361(92)90061-R).
- [150] Shafizadeh A, Shahbeig H, Nadian MH, Mobli H, Dowlati M, Gupta VK, et al. Machine learning predicts and optimizes hydrothermal liquefaction of biomass. *Chemical Engineering Journal* 2022;445. <https://doi.org/10.1016/j.cej.2022.136579>.
- [151] Ding J, Xu N, Nguyen MT, Qiao Q, Shi Y, He Y, et al. Machine learning for molecular thermodynamics. *Chin J Chem Eng* 2021;31:227–39. <https://doi.org/10.1016/J.CJCHE.2020.10.044>.
- [152] Keshavaraj R, Tock RW, Narayan RS, Bartsch RA, Ethridge D, Zhu R. Fluid property prediction of siloxanes with the aid of artificial neural nets. *Polymer - Plastics Technology and Engineering* 1996;35:971–82. <https://doi.org/10.1080/03602559608000611>.
- [153] Hall LH, Story CT. Boiling point and critical temperature of a heterogeneous data set: QSAR with atom type electrotopological state indices using artificial neural networks. *J*

- Chem Inf Comput Sci 1996;36:1004–14. <https://doi.org/10.1021/ci960375x>.
- [154] Jinwei Gao; Xueye Wang; Xinliang Yu; Xiaobing Li & Hanlu Wang. No Title: Calculation of polyamides melting point by quantum-chemical method and BP artificial neural networks. J Mol Model 2006;12:521–527. <https://doi.org/https://doi.org/10.1007/s00894-005-0087-6>.
- [155] Katritzky AR, Slavov SH, Dobchev DA, Karelson M. Rapid QSPR model development technique for prediction of vapor pressure of organic compounds. Comput Chem Eng 2007;31:1123–30. <https://doi.org/10.1016/j.compchemeng.2006.10.001>.
- [156] Abdul Jameel AG. Predicting sooting propensity of oxygenated fuels using artificial neural networks. Processes 2021;9:1070. <https://doi.org/10.3390/PR9061070/S1>.
- [157] Yan S, Eddings EG, Palotas AB, Pugmire RJ, Sarofim AF. Prediction of sooting tendency for hydrocarbon liquids in diffusion flames. Energy and Fuels 2005;19:2408–15. <https://doi.org/10.1021/EF050107D/ASSET/IMAGES/LARGE/EF050107DF00006.JPEG>.
- [158] Zhang C, Wang R, Hui X, Xue X, Lin Y, Sung CJ. Nonlinear Threshold Sooting Index Prediction Method for Surrogate Formulation Emulating Sooting Characteristics: A Case Study Using RP-3 Jet Fuels. Energy and Fuels 2020;34:9990–9. https://doi.org/10.1021/ACS.ENERGYFUELS.0C00921/ASSET/IMAGES/LARGE/EF0C00921_0009.JPEG.
- [159] Ye X, Arab P, Ahmadov R, James E, Grell GA, Pierce B, et al. Evaluation and intercomparison of wildfire smoke forecasts from multiple modeling systems for the 2019 Williams Flats fire. Atmos Chem Phys 2021;21:14427–69. <https://doi.org/10.5194/acp-21-14427-2021>.
- [160] Lawal R, Farooq W, Abdulraheem A, Abdul Jameel AG. Predicting soot formation in fossil fuels: A comparative study of regression and machine learning models. Digital Chemical Engineering 2024;12:100172. <https://doi.org/10.1016/J.DCHE.2024.100172>.
- [161] Li L, Sunderland PB. An Improved Method of Smoke Point Normalization. Combustion Science and Technology 2012;184:829–41. <https://doi.org/10.1080/00102202.2012.670333>.
- [162] Barrientos EJ, Anderson JE, Maricq MM, Boehman AL. Particulate matter indices using fuel smoke point for vehicle emissions with gasoline, ethanol blends, and butanol blends. Combust Flame 2016;167:308–19. <https://doi.org/10.1016/j.combustflame.2016.01.034>.
- [163] Haas FM, Qin A, Dryer FL. “Virtual” smoke point determination of alternative aviation kerosenes by threshold sooting index (TSI) methods. 50th AIAA/ASME/SAE/ASEE Joint Propulsion Conference 2014 2014:1–9. <https://doi.org/10.2514/6.2014-3468>.
- [164] Drews A. Standard Test Method for Smoke Point of Kerosine and Aviation Turbine Fuel. Manual on Hydrocarbon Analysis, 6th Edition 2008:263–263–6. <https://doi.org/10.1520/mnl10868m>.
- [165] Chang C-C, Chen S-H. A comparative analysis on artificial neural network-based two-stage clustering under a Creative Commons Attribution (CC-BY) 4.0 license A comparative analysis on artificial neural network-based two-stage clustering A comparative analysis on artificial neu 2015. <https://doi.org/10.1080/23311916.2014.995785>.
- [166] Mahmodi K, Mostafaei M, Mirzaee-Ghaleh E. Detecting the different blends of diesel and biodiesel fuels using electronic nose machine coupled ANN and RSM methods. Sustainable Energy Technologies and Assessments 2022;51:101914.

- <https://doi.org/10.1016/j.seta.2021.101914>.
- [167] Otchere DA, Arbi Ganat TO, Gholami R, Ridha S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *J Pet Sci Eng* 2021;200. <https://doi.org/10.1016/j.petrol.2020.108182>.
 - [168] Aydın M, Uslu S, Bahattin Çelik M. Performance and emission prediction of a compression ignition engine fueled with biodiesel-diesel blends: A combined application of ANN and RSM based optimization. *Fuel* 2020;269. <https://doi.org/10.1016/j.fuel.2020.117472>.
 - [169] Lamberg A, Toldy A, Keskiäli J, Karvo A, Larmi M, Santasalo-Aarnio A. Predicting Distillation Properties of Gasoline Fuel Blends using Machine Learning. *SAE Technical Papers* 2022. <https://doi.org/10.4271/2022-01-1086>.
 - [170] Kuzhagaliyeva N, Horváth S, Williams J, Nicolle A, Sarathy SM. Artificial intelligence-driven design of fuel mixtures. *Commun Chem* 2022;5:1–10. <https://doi.org/10.1038/s42004-022-00722-3>.
 - [171] Abdul Jameel AG, Oudenhoven VCO, Naser N, Emwas AH, Gao X, Sarathy SM. Predicting Ignition Quality of Oxygenated Fuels Using Artificial Neural Networks. *SAE Int J Fuels Lubr* 2021;14:4–14. <https://doi.org/10.4271/04-14-02-0005>.
 - [172] Giwa SO, Adekomaya SO, Adama KO, Mukaila MO. Prediction of selected biodiesel fuel properties using artificial neural network. *Frontiers in Energy* 2015;9:433–45. <https://doi.org/10.1007/s11708-015-0383-5>.
 - [173] Can Ö, Baklacioglu T, Öztürk E, Turan O. Artificial neural networks modeling of combustion parameters for a diesel engine fueled with biodiesel fuel. *Energy* 2022;247. <https://doi.org/10.1016/j.energy.2022.123473>.
 - [174] Prabhu AV, Alagumalai A, Jodat A. Artificial neural networks to predict the performance and emission parameters of a compression ignition engine fuelled with diesel and preheated biogas–air mixture. *J Therm Anal Calorim* 2021;145:1935–48. <https://doi.org/10.1007/S10973-021-10683-9/TABLES/7>.
 - [175] Tuan Hoang A, Nižetić S, Chyuan Ong H, Tarelko W, Viet Pham V, Hieu Le T, et al. A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels. *Sustainable Energy Technologies and Assessments* 2021;47. <https://doi.org/10.1016/J.SETA.2021.101416>.
 - [176] Chen C, Jiang X. Molecular dynamics simulation of soot formation during diesel combustion with oxygenated fuel addition. *Physical Chemistry Chemical Physics* 2020;22:20829–36. <https://doi.org/10.1039/D0CP01917H>.
 - [177] García A, Monsalve-Serrano J, Villalta D, Zubel M, Pischinger S. Potential of 1-octanol and di-n-butyl ether (DNBE) to improve the performance and reduce the emissions of a direct injected compression ignition diesel engine. *Volume 177, Pages 563 - 571* 2018;177:563–71. <https://doi.org/10.1016/j.enconman.2018.10.009>.
 - [178] Sarathy SM, Oßwald P, Hansen N, Kohse-Höinghaus K. Alcohol combustion chemistry. *Prog Energy Combust Sci* 2014;44:40–102. <https://doi.org/10.1016/J.PECS.2014.04.003>.
 - [179] Makwana A, Iyer S, Linevsky M, Santoro R, Litzinger T, O'Connor J. Effects of Fuel Molecular Weight on Emissions in a Jet Flame and a Model Gas Turbine Combustor. *J Eng Gas Turbine Power* 2018;140. <https://doi.org/10.1115/1.4037928/473633>.
 - [180] Liotta FJ, Montalvo DM. The Effect of Oxygenated Fuels on Emissions from a Modern Heavy-Duty Diesel Engine. *SAE Technical Papers* 1993.

- <https://doi.org/10.4271/932734>.
- [181] McEnally CS, Ciuparu DM, Pfefferle LD. Experimental study of fuel decomposition and hydrocarbon growth processes for practical fuel components: heptanes. *Combust Flame* 2003;134:339–53. [https://doi.org/10.1016/S0010-2180\(03\)00113-5](https://doi.org/10.1016/S0010-2180(03)00113-5).
 - [182] vom Lehn F, Cai L, Tripathi R, Broda R, Pitsch H. A property database of fuel compounds with emphasis on spark-ignition engine applications. *Applications in Energy and Combustion Science* 2021;5:100018. <https://doi.org/10.1016/J.JAECS.2020.100018>.
 - [183] Hoppe F, Burke U, Thewes M, Heufer A, Kremer F, Pischinger S. Tailor-Made Fuels from Biomass: Potentials of 2-butanone and 2-methylfuran in direct injection spark ignition engines. *Fuel* 2016;167:106–17. <https://doi.org/10.1016/J.FUEL.2015.11.039>.
 - [184] Westbrook CK. Biofuels combustion*. *Annu Rev Phys Chem* 2013;64:201–19. <https://doi.org/10.1146/ANNUREV-PHYSCHEM-040412-110009/CITE/REFWORKS>.
 - [185] Al Shebli MN, Raj A, Elkadi M, Anjum D, Pena GDJ, Prabhu A. Fuel oxygenation as a novel method to reduce sooting propensity of fuels: An investigation with gasoline surrogate fuels. *Fuel* 2022;324:124562. <https://doi.org/10.1016/j.fuel.2022.124562>.
 - [186] Uddin MT, Uddiny MA. Human activity recognition from wearable sensors using extremely randomized trees. *2nd International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2015* 2015. <https://doi.org/10.1109/ICEEICT.2015.7307384>.
 - [187] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat Comput* 2004;14:199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88/METRICS>.
 - [188] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 2011;12:2825–30.
 - [189] Awad M, Khanna R. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* 2015:1–248. <https://doi.org/10.1007/978-1-4302-5990-9/COVER>.
 - [190] Liu A, Sun JW, Ngo A, Mabadeje AO, Hernandez-Mejia JL. Machine Learning-Based Completions Sequencing for Well Performance Optimization 2024.
 - [191] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2016;13-17-August-2016:785–94. <https://doi.org/10.1145/2939672.2939785>.
 - [192] Machado MR, Karray S, De Sousa IT. LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. *14th International Conference on Computer Science and Education, ICCSE 2019* 2019:1111–6. <https://doi.org/10.1109/ICCSE.2019.8845529>.
 - [193] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. *Adv Neural Inf Process Syst* 2017;2017-December:4766–75.

7 Appendix

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor, ExtraTreesRegressor
from sklearn.svm import SVR, NuSVR
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
import lightgbm as lgb
import xgboost as xgb
import joblib

# Load the dataset
data = pd.read_excel('path_to_file.xlsx')
X = data.iloc[:, :-1]
y = data.iloc[:, -1]

# Normalize data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Split into training and testing
x_train, x_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# Define model hyperparameters for GridSearch
param_rf = {'n_estimators': [100, 200], 'max_depth': [5, 10], 'min_samples_split': [2, 5],
            'min_samples_leaf': [1, 2]}
param_svr = {'C': [0.1, 1, 10], 'kernel': ['linear'], 'gamma': ['scale']}
param_nusvr = {'C': [0.1, 1], 'nu': [0.5, 0.8], 'kernel': ['linear']}
param_lgb = {'learning_rate': [0.01, 0.1], 'n_estimators': [100, 200], 'max_depth': [5, 10]}
param_xgb = {'learning_rate': [0.1, 0.2], 'n_estimators': [100, 200], 'max_depth': [3, 5]}

# Train and tune Random Forest
rf = RandomForestRegressor(random_state=42)
grid_rf = GridSearchCV(rf, param_rf, cv=3, n_jobs=-1)
grid_rf.fit(x_train, y_train)
best_rf = grid_rf.best_estimator_

# Train and tune SVR
svr = SVR()
grid_svr = GridSearchCV(svr, param_svr, cv=3, n_jobs=-1)
grid_svr.fit(x_train, y_train)
best_svr = grid_svr.best_estimator_

# Train and tune NuSVR
nusvr = NuSVR()
grid_nusvr = GridSearchCV(nusvr, param_nusvr, cv=3, n_jobs=-1)
grid_nusvr.fit(x_train, y_train)
best_nusvr = grid_nusvr.best_estimator_

# Train and tune LightGBM
```

```

lgb_model = lgb.LGBMRegressor()
grid_lgb = GridSearchCV(lgb_model, param_lgb, cv=3, n_jobs=-1)
grid_lgb.fit(x_train, y_train)
best_lgb = grid_lgb.best_estimator_

# Train and tune XGBoost
xgb_model = xgb.XGBRegressor(objective='reg:squarederror')
grid_xgb = GridSearchCV(xgb_model, param_xgb, cv=3, n_jobs=-1)
grid_xgb.fit(x_train, y_train)
best_xgb = grid_xgb.best_estimator_

# Train and tune Extra Trees Regressor
param_etr = {'n_estimators': [100, 200], 'max_depth': [5, 10], 'min_samples_split': [2, 5],
'min_samples_leaf': [1, 2]}
etr = ExtraTreesRegressor(random_state=42)
grid_etr = GridSearchCV(etr, param_etr, cv=3, n_jobs=-1)
grid_etr.fit(x_train, y_train)
best_etr = grid_etr.best_estimator_

# Evaluate and save results
def evaluate_model(model, name):
    y_pred = model.predict(x_test)
    r2 = r2_score(y_test, y_pred)
    rmse = mean_squared_error(y_test, y_pred, squared=False)
    mae = mean_absolute_error(y_test, y_pred)
    print(f"{name}: R²={r2:.4f}, RMSE={rmse:.4f}, MAE={mae:.4f}")
    joblib.dump(model, f"{name}.pkl")

# Evaluate all models
evaluate_model(best_rf, "Random_Forest")
evaluate_model(best_svr, "SVR")
evaluate_model(best_nusvr, "NuSVR")
evaluate_model(best_lgb, "LightGBM")
evaluate_model(best_xgb, "XGBoost")
evaluate_model(best_etr, "Extra_Trees")

# Visualize predicted vs actual
def plot_results(actual, predicted, title):
    plt.scatter(actual, predicted, alpha=0.7)
    plt.plot([actual.min(), actual.max()], [actual.min(), actual.max()], 'r--')
    plt.xlabel("Actual")
    plt.ylabel("Predicted")
    plt.title(title)
    plt.show()

# Generate plots for the best model
for model, name in zip([best_rf, best_svr, best_nusvr, best_lgb, best_xgb, best_etr],
["Random_Forest", "SVR", "NuSVR", "LightGBM", "XGBoost", "Extra_Trees"]):
    plot_results(y_test, model.predict(x_test), f"{name}: Predicted vs Actual")
# Generate Taylor Diagram
observations = {'std': np.std(y_test)}
models = {

```

```

    "Random_Forest": {'std': np.std(best_rf.predict(x_test)), 'corr_coeff': np.corrcoef(y_test,
best_rf.predict(x_test))[0, 1], 'pbias': mean_squared_error(y_test, best_rf.predict(x_test),
squared=False)},
    "SVR": {'std': np.std(best_svr.predict(x_test)), 'corr_coeff': np.corrcoef(y_test,
best_svr.predict(x_test))[0, 1], 'pbias': mean_squared_error(y_test, best_svr.predict(x_test),
squared=False)},
    "NuSVR": {'std': np.std(best_nusvr.predict(x_test)), 'corr_coeff': np.corrcoef(y_test,
best_nusvr.predict(x_test))[0, 1], 'pbias': mean_squared_error(y_test, best_nusvr.predict(x_test),
squared=False)},
    "LightGBM": {'std': np.std(best_lgb.predict(x_test)), 'corr_coeff': np.corrcoef(y_test,
best_lgb.predict(x_test))[0, 1], 'pbias': mean_squared_error(y_test, best_lgb.predict(x_test),
squared=False)},
    "XGBoost": {'std': np.std(best_xgb.predict(x_test)), 'corr_coeff': np.corrcoef(y_test,
best_xgb.predict(x_test))[0, 1], 'pbias': mean_squared_error(y_test, best_xgb.predict(x_test),
squared=False)},
    "Extra_Trees": {'std': np.std(best_etr.predict(x_test)), 'corr_coeff': np.corrcoef(y_test,
best_etr.predict(x_test))[0, 1], 'pbias': mean_squared_error(y_test, best_etr.predict(x_test),
squared=False)}
}

_ = taylor_plot(
    observations, simulations=models,
    title="Taylor Diagram",
    leg_kws={"facecolor": "white", "edgecolor": "black", "fontsize": 12},
    marker_kws={"ms": 8, "markeredgecolor": "k"},
    cont_kws={"colors": "blue", "linestyles": "--"},
    grid_kws={"color": "gray"}
)

```

ANN model

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau, ModelCheckpoint

# Load data
path = "C:/Users/DELL/Downloads/paper for SP and soot/TPRFdata - withoutO-N.xlsx"
df_raw = pd.read_excel(path).drop("Name", axis=1)

# Preprocessing
X = df_raw.iloc[:, :-1].to_numpy()
Y = df_raw.iloc[:, -1].to_numpy()

scaler = StandardScaler()

```

```

X_scaled = scaler.fit_transform(X)
Y_scaled = scaler.fit_transform(Y.reshape(-1, 1)).flatten()

# Splitting the data
X_train, X_temp, y_train, y_temp = train_test_split(X_scaled, Y_scaled, test_size=0.3,
random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, random_state=42)

# Build the model
def create_model():
    model = Sequential([
        Dense(30, input_dim=X.shape[1], activation='tanh',
kernel_regularizer=tf.keras.regularizers.l2(0.01)),
        Dropout(0.3),
        Dense(70, activation='tanh', kernel_regularizer=tf.keras.regularizers.l2(0.01)),
        Dropout(0.3),
        Dense(50, activation='tanh', kernel_regularizer=tf.keras.regularizers.l2(0.01)),
        Dropout(0.3),
        Dense(1, activation='linear', kernel_regularizer=tf.keras.regularizers.l2(0.01))
    ])
    model.compile(optimizer=Adam(learning_rate=0.001), loss='mean_squared_error',
metrics=['mean_squared_error'])
    return model

# Callbacks
checkpoint_filepath = 'best_model_weights.h5'
callbacks = [
    ModelCheckpoint(filepath=checkpoint_filepath, save_best_only=True, verbose=1),
    EarlyStopping(monitor='val_loss', patience=20, restore_best_weights=True),
    ReduceLROnPlateau(monitor='val_loss', factor=0.1, patience=5, min_lr=1e-6, verbose=1)
]

# Train the model
model = create_model()
history = model.fit(X_train, y_train, validation_data=(X_val, y_val), batch_size=32, epochs=600,
callbacks=callbacks)

# Evaluate on test data
y_test_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_test_pred)
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_test_pred)
print(f"Test RMSE: {rmse}, R²: {r2}")

# Plot learning curve
plt.figure()
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title('Learning Curve')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()

```

```
# Predicted vs Actual
plt.scatter(y_test, y_test_pred, alpha=0.6)
plt.plot([min(y_test), max(y_test)], [min(y_test), max(y_test)], 'r--')
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Predicted vs Actual (Test Data)')
plt.show()

# Save the model
model.save("TPRF_SPMModel_V2.h5")
```

Vitae

Name	Ruaa Ala Eldin Ageeb Abakar
Nationality	Sudanese
Email	ruaa_abakar@hotmail.com
Adress	Gamma Ave, Chemical Engineering Department, University of Khartoum, P.O. Box 321, Postal/Zip Code 11115, Khartoum, Sudan
Academic Background	Master Of Science in Chemical Engineering King Fahd University of Petroleum and Minerals (First Class, CGPA:4.0 out of 4.0) 2022-2024
Publication	Abakar. R, Abul Jameel. A “Impact of Blending Alcohols, Ethers and Ketones with TPRF MIXTURES on Smoke Point. An Experimental and Neural Network Analysis” (In preparation)
Conference Presentation	“Predicting Smoke Points of Oxygenated Blended Fuels: Leveraging Artificial Neural Networks for Enhanced Soot Reduction and Renewable

Energy Solutions" in the Artificial Intelligence in
Materials and Process Design track

Canadian Society For Chemical Engineering
(CSCHE)2024, Toronto, Canada

ACS/RSC Regional MENA Conference 2024,
Doha, Qatar

Achievement

Award:1st Place CHEMLLMATHON,
Chemindex 2024, Dhahran, Saudi Arabia
November 2024

Award: Best Poster Presentation, Canadian
Society for Chemical Engineering (CSCHE)2024,
Toronto, Canada

Award: Renewable Energy Technical Incubator
(RETI) Graduate Research Grant Program,
Dhahran, KSA

Award: KFUPM Scholarship Recipient, Ms. In
Chemical Engineering August 2022

Best Academic Excellence Award, University of
Khartoum

Certificate of Distinction, State of Khartoum

Honor Student Prize Ceremony, Riyadh, KSA

Skills

1. Proficient in machine learning and deep learning models.
2. Expertise in Python, MATLAB and integrated development environments (IDEs) like Jupyter and Visual Studio.
3. Expertise in measuring density, flash point, surface tension, smoke point, viscosity and octane rating.
4. Expertise in Characterization techniques including XRD, TGA, MicroGC, SEM and FT-IR
5. Excellent in data analysis and scientific writing.
6. Strong presentation and communication skills.