OPTIMIZATION OF ENERGY DENSITY IN SUPERCAPACITORS BY UTILIZING A HYBRID ARTIFICIAL NEURAL NETWORKS-GENETIC ALGORITHM BASED OPTIMIZATION ALGORITHM

by

Duygu Kaya

B.S., Chemical Engineering, Yıldız Technical University, 2019

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2022

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor Assist. Prof. Betül Uralcan for her endless support, guidance, patience, and interest in every detail of my research. I am also very grateful to my co-supervisor Assist. Prof. Erdal Aydın who was encouraged and helped me in every situation. I owe my deepest gratitude to both that makes this period easy and enjoyable for me.

I would also like to thank Prof. Ramazan Yıldırım and Assist. Prof. Hasan Şıldır for their valuable time, comments, and acceptance to be on my thesis committee.

I am very grateful to the Soft Matter Laboratory group for their support and encouragement. Also, I would like to thank Dilara Saadetnejad and Doğan Tarık Karabay for enabling me to use the dataset in my thesis and special thanks to Dilara for answering my questions at any time and for valuable friendship. I am also very thankful to Process Systems Engineering Lab for their support, friendship and help.

I am sincerely grateful to Abdullah Aladağ for his precious friendship, understanding, and help for many years. Also, I would like to express my appreciation to my lab mate Ayşe Saliha Korkut for her intimacy, kindness, and being with me at every step of this period. I would also like to thank Akarun Ayça Ersoy, Hakan Sabri Sakallıoğlu, and İlke Seymen Keykanat for their endless support and motivation. It was very meaningful to spend time with you during the difficult times of this research.

I would like to extend my thanks to Özge Selçuk, Sercan Altundemir, Esra Ülgey, Pınar Eribol, Özlem Özbek for their valuable friendship, support, and motivation.

I would also like to thank all of my professors for their efforts, guidance, and insightful comments.

Also, special thanks to my all friends who motivating and encouring me. Thank you very much for your unconditional support, interest, and cheerful conversations, being your friend is a great chance for me.

Finally, I would like to express my deepest gratitude to my family for their love and support. I am very lucky to have such an amazing family. Thanks to my parents for their continuous encouragement throughout my life. Also, I am especially grateful to my brother Doğuş Kaya for being there for me whenever I needed him and making me laugh.

This work was supported by grants from The Scientific and Technological Research Council of Turkey (TUBITAK) under BIDEB 2232-A International Fellowship for Outstanding Researchers (Project Number T118C220) and Bogazici University Research Fund (Project Number 17841).

ABSTRACT

OPTIMIZATION OF ENERGY DENSITY IN SUPERCAPACITORS BY UTILIZING A HYBRID ARTIFICIAL NEURAL NETWORKS-GENETIC ALGORITHM BASED OPTIMIZATION ALGORITHM

Energy storage systems are required to meet the increased energy demand and reduce the need for fossil fuels. Energy storage devices are of interest to renewable energy systems and electric vehicles to provide a permanent energy supply. Supercapacitors, in particular, are eligible systems for energy storage owing to their unique properties such as very long cycle life, high reversibility, and high power density. Nonetheless, they have limited energy density. The ultimate goal is to increase the energy density of supercapacitors while maintaining high power density. In this thesis, hybrid artificial neural network-genetic algorithm (ANN-GA) model is utilized to increase the capacitance of supercapacitors. Several data preprocessing, feature selection, and machine learning algorithms are performed to predict the capacitance of supercapacitor by using experimental data. It is observed that ANN is a powerful method to capture nonlinear relationships concerning the physical and operational features of supercapacitors. GA is a promising method that examines search space for the optimal solution. The trained neural network model is used as the fitness function for genetic algorithm to achieve maximum capacitance within the feasible range. Selection, crossover, and mutation procedures are implemented in the reproduction step of GA to offer elaborate search space. In a nut shell, this study takes a step towards the rational design of supercapacitors by implementing a hybrid ANN-GA as an optimization tool to improve the capacitance. The results indicate that obtained optimal design parameters agree with the literature while improving the capacitance of supercapacitors significantly.

ÖZET

HİBRİT YAPAY SİNİR AĞLARI-GENETİK ALGORİTMA MODELİ KULLANARAK SÜPERKAPASİTÖRLERİN ENERJİ YOĞUNLUĞUNUN OPTİMİZE EDİLMESİ

Enerji depolama sistemleri artan enerji ihtiyacını karşılamak ve fosil kaynaklara olan bağlılığı azaltmak için gereklidir. Yenilenebilir enerji sistemleri ve elektrikli araçlar için uzun süreli enerji tedarik imkânı sunması açısından ilgi çekici sistemlerdir. Süperkapasitörler, çok uzun çevrim ömrü, hızlı şarj/deşarj döngüsü ve güç yoğunluğu gibi benzersiz özelliklerinden dolayı enerji depolamaya elverişli sistemlerdir, ancak enerji yoğunlukları düşüktür. Güç yoğunluğunu sabit tutarken enerji yoğunluğunu arttırmak süperkapasitör için temel hedeftir. Bu çalışmada, süperkapasitörlerin kapasitansını artırmak için hibrit Yapay Sinir Ağı-Genetik Algoritma modeli kullanılmıştır. Süperkapasitör kapasitansını tahmin etmek için deneysel çalışmalardan toplanan veriyi kullanarak çeşitli veri ön işleme, özellik seçimi ve makine öğrenmesi algoritmaları kullanılmıştır. Yapay sinir ağlarının süperkapasitörlerin fiziksel ve çalışma şartları ile ilgili doğrusal olmayan ilişkileri yakalamak için güçlü bir yöntem olduğu gözlemlenmiştir. Genetik algoritma, optimal çözüm için arama uzayını inceleyen etkileyici bir yöntemdir. Eğitilmiş sinir ağı modeli, uygulanabilir limitler içinde optimum kapasitans elde etmek için genetik algoritmanın uygunluk fonksiyonu olarak kullanılmıştır. Seçme, çaprazlama ve mutasyon gibi genetik algoritma prosedürleri arama uzayını artırmak için uygulanmıştır. Bu çalışma, kapasitansı iyileştirmek için bir optimizasyon aracı olarak hibrit sinir ağı-genetik algoritması kullanarak süper kapasitörlerin rasyonel tasarımını önermiştir. Sonuçlar, literatür taramasına göre optimum çözümün mümkün olduğunu göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a	Step Size
A	Surface area of electrode
b	Binary variable
B	Total number of trees
C	Capacitance
c	Constant response
D	Number of inputs
E	Energy
e_i	Prediction errors
$\mathbf{g}$	Gradient
$\mathbf{H}$	Hessian matrix
$\mathbf{I}$	Identity matrix
$\mathbf{J}$	Jacobian matrix
p_c	Crossover probability
p_m	Mutation probability
P_{max}	Maximum instantaneous power
R	Internal resistance
R^2	Coefficient of determination
V	Voltage
Q	Stored charge
w_0	Bias of artificial neural networks
w_i	Weights of artificial neural networks
β_o	Intercept of linear regression
β_j	Coefficient for regression
ϵ_o	Permittivity
ϵ_r	Dielectric constant of insulation material
Θ_b	b^{th} random forest tree

λ	Shrinkage parameter
μ	Combination coefficient

LIST OF ACRONYMS/ABBREVIATIONS

3D	Three Dimensional
ANN	Artificial Neural Networks
APW	Absolute Potential Window
BO	Bayesian Optimization
CN	Carbon Nanosphere
CNT	Carbon Nanotube
CV	Cross-Validation
CVD	Chemical Vapor Deposition
DWCNT	Double-Walled Carbon Nanotube
EA	Evolutionary Algorithms
EBP	Error Backpropagation
GA	Genetic Algorithm
GWO	Gray Wolf Optimization
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MINLP	Mixed Integer Nonlinear Programming
ML	Machine Learning
MLP	Multilayer Perceptron
NSR	Normalized Scan Rate
OOB	Out-of-Bag
PBI	Polybenzimidazole
PCM	Porous Carbon Materials
PSO	Particle Swarm Optimization
PTFE	Polytetrafluoroethylene
PVDF	Polyvinylidene Fluoride
RF	Random Forest
RMSE	Root Mean Square Error
RSS	Residual Sum of Squares

SA	Sensitivity Analysis
SSA	Specific Surface Area
SSE	Sum of Squares Error
SVM	Support Vector Machine
SQP	Sequential Quadratic Programming

1. INTRODUCTION

Renewable energy systems are efficient, sustainable, clean and eco-friendly energy sources that can satisfy increasing global energy demand. The seasonal impact is an important parameter for many renewable energy sources such as wind and solar systems. Hence, promising energy storage systems are needed instead of conventional storage technologies to provide advancement of renewable-based energy [1].

Electric vehicles, home energy systems, and renewable energy systems need energy storage devices such as supercapacitors and batteries [2]. Energy and power density, speed of charging, size, weight, and cost are the main requirements for energy storage equipment. Lithium-ion batteries (LIBs) with high energy, power density, low cost, and cycle life are satisfactory energy storage devices except for the ability of fast charging. In spite of these benefits, the main drawback of LIBs is potential fire due to the increased heat generation. Also, rapid charging results in enhanced degradation of the battery [3,4]. Supercapacitors, also known as electric double-layer capacitors (EDLCs), store the charge electrostatically on active materials that have high specific surface and electrochemical stability by using reversible adsorption of electrolyte ions [5]. Supercapacitors have received more attention in recent years due to the higher charge/discharge efficiency, life cycle, and power density. These characteristics have attracted significant attention since they facilitate building more advanced energy storage systems, for onboard and future applications. However, a major limitation of supercapacitors is low energy density [6]. The performance evaluation of energy storage devices is shown in Figure 1.1 which is called Ragone plot [7].

Supercapacitors, also known as EDLCs or ultracapacitors, utilize high surface area electrode materials and achieve larger capacitances than conventional capacitors by storing electrical charges in a thin double layer at the electrode/electrolyte interface. Supercapacitors, similar to conventional capacitors, rely on energy storage through a physical mechanism. However, they integrate electrodes with much higher surface areas

and much thinner dielectrics to decrease the distance between the electrodes. Some of the features such as specific density, discharge rate and cycle life of electrochemical storage devices are compared in Table 1.1.

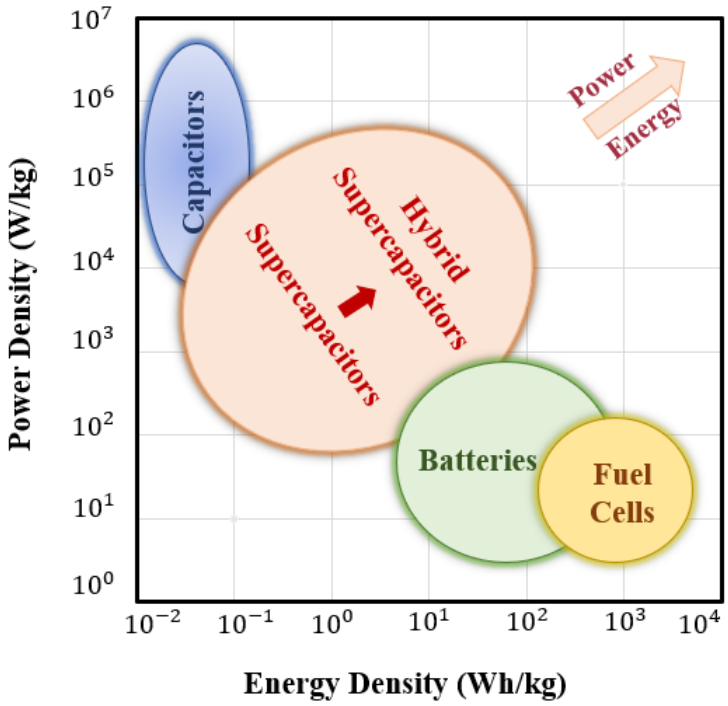

Figure 1.1. Ragone Plot [7].

Conventional capacitors include two conducting electrodes and an insulating dielectric material which separates electrodes. Opposite charges are attracted to the surfaces of each electrode when a voltage is applied to a capacitor. These charges are kept separate by dielectric to produce an electric field that provides energy storage for capacitor. Capacitance formula can be represented as,

$$C = \frac{Q}{V} \tag{1.1}$$

where C is capacitance, Q is stored charge and V is voltage.

Table 1.1. Comparison of electrochemical storage devices [24].

Characteristics	**Capacitor**	**Supercapacitor**	**Battery**
Specific energy ($Whkg^{-1}$)	<0.1	1-10	10-100
Specific power (Wkg^{-1})	>10000	500-10000	<1000
Discharge time	10^{-6} to 10^{-3}	s to min	0.3-3 h
Charge time	10^{-6} to 10^{-3}	s to min	1-5 h
Coulombic efficiency (%)	About 100	85-98	70-85
Charge time	Almost infinite	>500000	About 1000

For conventional capacitance, the capacitance can be expressed as,

$$C = \varepsilon_o \varepsilon_r \frac{A}{D} \tag{1.2}$$

since the capacitance is proportional to surface area of electrodes and inversely proportional to distance between electrodes where ε_o is permittivity and ε_r is the dielectric constant of insulating material. Energy density and power density are two primary performance criteria of capacitors. Energy storage is proportional to capacitance of capacitor that can be represented as,

$$E = \frac{1}{2} C V^2. \tag{1.3}$$

The maximum instantaneous power P_{max} can be defined as,

$$P_{max} = \frac{V^2}{4R} \tag{1.4}$$

where R is internal resistance of capacitor. Supercapacitors have similar attributes to conventional capacitors. However, they have higher surface area and lower distance between electrodes. When voltage is applied, charge accumulates on the electrode surface. Then, the oppositely charged ions in the solution are attracted to electrode surface that can be seen in Figure 1.2. Since there is no chemical reaction, charge

storage is highly reversible which provides high cycling stability [8].

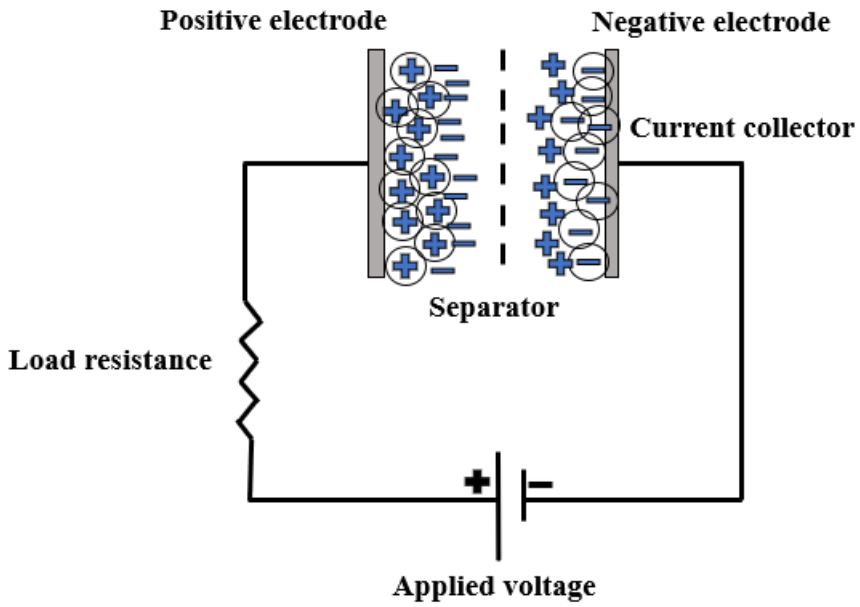

Figure 1.2. Schematic of supercapacitors.

Capacitance, equivalent series resistance and operating voltage are essential parameters to evaluate energy and power performance of supercapacitors [6]. Electrode material, electrolyte, current collector, and separator are basic components that play an important role in supercapacitors that can be seen in Figure 1.2. Electrode is the essential component that affects performance of supercapacitors since the type of electrode-active material influence the charge storage. Large surface area, good redox activity and high electrical conductivity are desired characteristics for electrode materials [9]. Therefore, the key factor to achieve high capacitance is utilizing electrically conductive and high specific surface area electrodes. Carbon materials are the most common electrode materials thanks to the high surface area, low cost, abundance, and high conductivity. High specific surface area with numerous available active sites for interactions of ions has great importance on capacitance. However, there is no linear relationship between capacitance and specific surface area. It has been discovered that pore size and shape also affect capacitance. The pore size is a considerable parameter that provides the proper transfer of electrolyte ions and efficient interaction of

electrolyte ions with pores [10, 11].

Furthermore, electrical conductivity of electrode materials depends on their crystallization degree. The structure of crystallinity can be examined by using Raman spectroscopy that detects graphitic structures of carbon materials. The D-band represents the amorphous carbon and defects and G-band specifies the presence of sp^2 hybridized carbon atoms. The ratio between D-band and G-band demonstrates the degree of graphitization. A higher ID/IG ratio possesses a higher graphitic structure that ensures high electrical conductivity [9, 11].

Current collector is a significant parameter from various aspects for supercapacitor performance. Current collector supports electrode and attracts electrons from the electrolyte to transfer into the external circuit. Also, it should not react with electrolyte and electrode materials during charge/discharge procedure. Metal foams and foils are widely used in supercapacitors such as nickel foam and stainless steel. The active electrode material is usually mixed with a binder to obtain a continuous electrode film and decrease contact resistance between the current collector and electrode. Polytetrafluoroethylene (PTFE), polyvinylidene fluoride (PVDF) and polybenzimidazole (PBI) are widespread binder materials. Although separator do not play important role in the electrochemical reactions during energy storage, it is main component of supercapacitors. Separator is used to prevent short circuit between positive and negative electrodes of supercapacitors. For high performance supercapacitors, separator should be thin, high permeability, non-flammable, thermally stable and affordable. Fibers, polymers and natural substrates such as cellulose, rubber, polypropylene are commonly used separator materials in supercapacitors [9].

Voltage is also an important parameter since the main interest is increasing the energy density of supercapacitors. The relationship between energy density and voltage can be observed in Equation (1.3). However, the decomposition of electrolytes at high potentials hinders reaching the desirable wide voltage window. Aqueous electrolytes have low electrochemical stability as the stable operating potential window is 1.23 V.

Hence, organic electrolytes such as propylene carbonate or acetonitrile have become popular owing to the high voltage window (2.5-2.7 V) for supercapacitors. Ionic liquids are also attractive electrolyte materials that can enable higher than 4 V for voltage window. The design of electrolytes that can satisfy high voltage is necessary for energy storage applications [10]. Cyclic voltammetry method is used to measure capacitance of supercapacitors by applying electric potential between positive and negative electrodes. Scan rate refers the speed of the potential change. Lower scan rate leads higher capacitances in supercapacitor applications [9].

Moreover, the integration of heteroatoms such as boron, nitrogen, phosphorous into carbon materials have been implemented as an alternative method to increase capacitance. Nitrogen has received attention in recent years due to the pseudocapacitance effect and improving electrical conductivity of supercapacitors [12].

The main challenge about supercapacitors is increasing their energy storage capacity to provide high-performance energy storage systems for electrical devices and intermittent sources. Although high capacitance values may achieve by using several methods such as heteroatom-doped electrodes, low cell voltage of aqueous electrolytes (less than 1.5 V) due to the electrolysis of water limits the usage of aqueous solutions. Recently, commercial supercapacitors have combined with organic solvents and ionic liquids that may achieve higher cell voltage instead of aqueous electrolytes. Electrolyte materials that can tolerate high cell voltage improve the capacitance of supercapacitors drastically, yet high surface area active material with enhanced electrochemical stability is required to achieve desired energy storage capacity [5]. Therefore, favorable electrode/electrolyte materials, operating conditions, and design parameters that affect the capacitance of supercapacitors should be optimized to acquire an efficient energy storage device. In order to obtain the feasible set of parameters that gives optimum capacitance value, experimental studies may be carried out by considering several combinations of features. However, it can take a long time since several parameters affect the performance of supercapacitors. Also, it can be compelling to follow a pattern between relationships of features. Hence, computational models can be preferable to

solve such a problem due to the time constraint and practicality.

Recently, mechanism-based solutions have been replaced by machine and statistical learning methods with the advancement of computer technology. Machine learning is the process to solve specific problems by collecting a dataset and building algorithms based on statistical models. Machine learning algorithms such as artificial neural networks (ANN) become more and more popular for non-linear systems. Due to the non-linear characteristics of supercapacitors, ANN can be used to build a model and to investigate several problems [13, 14].

Supercapacitors have several advantages such as high-power density and long-cycle life. The electrode is one of the key components and carbon is the most widely used material as an electrode. Therefore, the specific capacitance of carbon-based electrodes is a significant criterion to enhance the performance of supercapacitors. Zhu *et al.* propose that using artificial neural networks to predict the capacitance performance of supercapacitors since ANN is a powerful estimation technique to examine non-linear behaviors. The dataset is generated from published papers about carbon-based supercapacitors. Specific surface area, pore size, ID/IG ratio, N-doping level, and voltage window are used to investigate their impact on capacitance. Also, lasso and linear regression methods are computed to predict capacitance to compare the performance of ANN. It is observed that ANN shows more accurate results compared to lasso and linear regression according to test results. Hence, this study implies that ANN can be an effective method in materials selection for the performance of supercapacitors [15].

Another study to predict the capacitance of carbon-based supercapacitors using machine learning is proposed by Su *et al.* In this work, seven features (specific surface area, pore size, pore volume, ID/IG ratio, N-doping percentage, O-doping percentage, and potential window) collected from experimental published papers are used to train machine learning algorithms. Linear regression, support vector machine, multilayer perceptron (MLP), and regression tree models are performed, and results are compared. Multilayer perceptron (MLP) is a commonly used artificial neural network model, which

consists of the input layer, hidden layers, and an output layer. Each node includes weights to connect all nodes in the next layer. The correlation coefficient (R), mean absolute error (MAE), and root mean square error (RMSE) were used as performance metrics to evaluate the accuracies of ML algorithms in predicting capacitance. It is obtained that the regression tree shows better performance based on RMSE results. However, there is a slight difference between the accuracies of the regression tree and the MLP model [16].

Structural features of carbon-based materials are necessary parameters to maintain the objective design of EDL materials. Gheytanzadeh *et al.* propose a support vector machine (SVM) by a grey wolf optimization (GWO) to investigate the correlation between structural features and capacitance. Specific surface area, calculated pore size, ID/IG ratio, N-doping level, and potential window are chosen to examine the impact of structural features on supercapacitors. The related parameters are extracted from experimental papers. The proposed SVM-GWO model indicates great accuracy with selected features and corresponding capacitances. Also, sensitivity analysis was performed, and it shows that specific surface area and pore volume have considerable influence on capacitance of supercapacitors [11].

Porous carbon materials (PCMs) are attractive materials as electrodes thanks to their beneficial features such as low cost, chemical stability, high surface area, and high conductivity. Another study is conducted by Liu *et al.* to examine the relationship between structural features of carbon and capacitance by using machine learning. Specific surface area is regarded as the most important factor that affects specific capacitance. However, it is reported that pore size and structure have an impact on specific capacitance according to the dimension of ionic species. In the present study, porous structural features are taken into consideration such as pore size, pore volume, specific surface area, surface area of micropores, pore volume of mesopores, etc. The data points are collected from experimental studies that are related to PCMs. In order to predict capacitance, various machine learning algorithms that consist of extreme gradient boosting, gradient boosting machines, multiple linear regression, ANN and

random forest are computed. It is observed that extreme gradient boosting is the best model to capture the correlation of structural features and capacitance [17].

Another study focused on hyperparameter tuning of machine learning algorithms while investigating structural features of carbon-based supercapacitors. Ahmed *et al.* propose fine-tuned artificial neural networks and random forest (RF) algorithms to predict specific capacitance, power density, and energy density. Inputs are scan rate, specific surface area, and micropores surface area. The inputs and outputs are extracted from research papers that include a three-electrode system in 6 M KOH electrolytes with 1 V potential window. In machine learning applications, hyperparameters that specify the learning procedure and complexity of the model are fixed before the training procedure. Thus, hyperparameter tuning is inevitable to achieve high accuracies. Activation function, number of neurons and hidden layers, optimization algorithm, Kernel initializer, number of batch and epoch for ANN; and number of trees for RF are tuned. Also, random forest is used to assess the importance of features. The results shows that hyperparameter tuning affects the accuracy of models to predict capacitance [18].

Farsi *et al.* propose that supercapacitor performance simulator by using artificial neural networks to avoid computation of partial differential equations which describe the system. The data is obtained from the model proposed by Lin *et al.* due to the lack of experimental data of supercapacitor development. Power density, energy density and utilization were simulated by using active materials lattice length, particle size, exchange current density, and the current density drawn from supercapacitors. Feed forward back propagation algorithm with Levenberg– Marquardt method is employed. It is observed that four-layer neural net with two hidden layers exhibited better results to simulate model. The supercapacitor performance predictions of ANN have very good correspondence with numerical method of Lin model [19].

Ding *et al.* propose that Support Vector Machine-Particle Swarm Optimization (SVM-PSO) model to predict output voltage of supercapacitors. The experimental set-up is built, and data collected to test effectiveness of the model. Support vector

machine is performed to predict output voltage of supercapacitors with temperature, current and initial voltage as an input. Particle swarm optimization is used to tune parameters of SVM that are penalty factor and width parameter. Penalty factor provides balance between complexity of the model and training error, width parameter shows the training sample characteristics. When these parameters are selected arbitrarily, the model can lead suboptimal solutions. Therefore, PSO algorithm is implemented to find two parameters since the performance of SVM for regression depends on the parameter selection. The results implies that SVM-PSO algorithm is efficient model to predict output voltage of supercapacitors [20].

Genetic algorithm (GA) is proposed as one of the first population-based stochastic algorithms. GA is based on Darwin's theory of evolution in which fittest is the survival. GA works on a population consisting of some solutions using fitness function. In this case, every solution represents chromosome. The chromosome is formed set of genes (features) that defines the individual. Selection mechanism such as roulette wheel or tournament selection is used in order to select the best individuals by using fitness function. The higher fitness value represents the higher the quality solution. Despite stochastic process of genetic algorithm, it is reliable method since the results are improved generation by generation and approximate the optimum. Also, crossing procedure is performed by using selected two solution (parent solution) that produce two new children solutions. Mutation operator changes the genes of chromosomes randomly to increase input space [21].

Zhou *et al.* suggest hybrid genetic algorithm to tune parameters of recurrent neural network model. In this paper, Genetic Algorithm-Sequential Quadratic Programming (GA-SQP) is applied to optimize hidden layer numbers and dropout probability of long short-term memory (LSTM) neural network. Firstly, aging tests of supercapacitors were examined for dynamic and steady-state profile to obtain dataset. Then, GA-SQP hybrid model is applied to improve structure of LSTM. Sequential quadratic programming is used to enhance the search ability of GA and provides fast convergence. Number of hidden layer and dropout probability of LSTM is optimized using GA-SQP

and these results are used to train model. Trained neural network model is implemented to predict remaining useful time of supercapacitors. It is observed that proposed GA-SQP algorithm effectively optimized parameters of LSTM model and neural network achieved high accuracies to predict remaining useful time of supercapacitors [22].

Machine learning can be useful to predict the performance of new materials for specific systems. The electrode material is a fundamental part of supercapacitors. Carbon-based materials are widely used as electrodes in supercapacitor applications, yet, there is a deficiency of research on metal oxynitrides. Metal oxynitrides enable unique properties for supercapacitors such as improved chemical inactivity and thermal stability. Also, nitrogen with the oxide phase provides rapid redox reactions that lead to increased capacitance. Ghosh *et al.* propose a data-driven approach to predict capacitance and cyclic stability of novel electrode material which is cerium oxynitride. The database includes metal oxides, metal nitrides, and metal oxynitrides such as Vanadium, Titanium, Tungsten, Cerium, etc. The dataset is collected from experimental studies for these materials that include specific surface area, voltage window, current density, current collector, and morphology. Rod, tube, spherical can be given as an example for morphology categories. Random Forest (RF) and Multilayer Perceptron (MLP) are used to build models. The predicted values of specific capacitance and cyclic retention for cerium oxynitride are 214 F/g (it is equivalent to 28.5 mAh/g) at 2 A/g current density and 90–100% capacitance retention with 10,000 cycles, respectively. The experimental procedure is carried out under similar conditions and the cerium oxynitride is synthesized via the urea glass method. It is achieved that 26 mAh/g for specific capacity and 100% capacity retention with cerium oxynitride under the certain experimental conditions that are used in prediction. There is no clear difference between experimental and predicted values. Therefore, this study emphasizes that machine learning provides an advantage to predict the performance of novel materials by using experimental validation [23].

Machine learning models have been used in several aspects such as prediction of capacitance, ageing and output voltage of supercapacitors. In aforementioned studies,

artificial neural networks are widely used machine learning algorithm to make predictions since it is useful method to handle complex and non-linear systems. Also, genetic algorithm is integrated into the system to optimize neural networks for supercapacitor domain. Although the usage of artificial neural networks and genetic algorithm have been studied in supercapacitors from various perspectives, there is a lack of investigation on hybrid Artificial Neural Network-Genetic Algorithm (ANN-GA) model.

Supercapacitors are devices that capable of supplying high power rates. However, their energy storage capacity is limited. For this reason, supercapacitors are suitable for applications where the high power is required but high energy storage capacity is not needed [24]. Therefore, supercapacitors can be improved by increasing energy storage capacity to provide effective energy storage systems.

The aim of this study is optimizing the energy storage capacity of supercapacitors to obtain highly efficient energy storage devices by utilizing a hybrid artificial neural network-genetic algorithm model. In that context, hybrid ANN-GA based optimization algorithm is implemented to find parameter sets that yield optimized capacitance. For this purpose, to the best of our knowledge, ANN-GA hybrid model is utilized for capacitance prediction for the first time. The data points are extracted from 110 published experimental articles and includes both categorical and numerical features. Several preprocessing and feature selection methods is implemented to eliminate redundant instances and features. Outlier detection, normalization, and feature selection methods such as random forest variable importance, trial/error method and creating subset of features are performed to get simple and efficient model.

Machine learning (ML) methods such as linear/lasso regression, tree-based ensembles are applied to compare performance of neural network to predict the capacitance of supercapacitors. Artificial neural network model is used to predict capacitance of supercapacitors by using the dataset that consists of features including specific surface area, electrolyte type, potential window, etc. Trained neural network model is used as the fitness function of genetic algorithm to maximize capacitance of supercapaci-

tors. Genetic algorithm is used to increase input space by using selection, crossover, and mutation procedures. Furthermore, features that give maximum capacitance is obtained as an output of genetic algorithm. The results are compared with literature information to assess their feasibility.

The thesis outline can be summarized as follows: In Section 2, statistical learning methods, genetic algorithm and proposed hybrid ANN-GA model are described. In Section 3, results of feature selection, different machine learning algorithms and optimal solution are reported. Finally, conclusion and outlook are presented in Section 4.

2. MATERIALS AND METHODS

2.1. Statistical Learning

2.1.1. Data Preprocessing

Data analysis is fundamental to solve problems and make decisions on specific topic from engineering to social sciences. The objective of data analysis is acquiring better insight into interested topic and discover basic knowledge. However, the reliability of data is substantial factor to obtain accurate and precise outcomes. Hence, data preprocessing is important step to eliminate potential problems that can be arise from data source and procedure of data collection. Data preprocessing comprise of all procedures before the implementation of actual processes to enhance effectiveness of dataset. Proper preprocessing enables to achieve more meaningful knowledge and pattern from the dataset [25]. Data preprocessing encompasses all data preparation and data reduction steps such as data cleaning, data integration, data transformation, normalization, and feature selection. Data preprocessing steps in machine learning is demonstrated in Figure 2.1. Data cleaning is used for handling inconsistent data and correcting wrong data. Usually, the raw data collected from several sources are not convenient in machine learning since they include various ranges and units. Normalization is required to express the features in same units and range [26].

Min-max normalization that shows a linear transformation on the original data is represented as

$$x_{normalized} = \frac{x - min(x)}{max(x) - min(x)}(new_{max} - new_{min}) + new_{min} \tag{2.1}$$

where x is the feature, $max(x)$ and $min(x)$ is the minimum and maximum values of feature and new_{max} and new_{min} are desired range of the values.

Denormalization formula is presented as

$$x_{real} = x_{normalized}\frac{max(x) - min(x)}{(new_{max} - new_{min})} + min(x) - new_{min} \tag{2.2}$$

where new_{max} and new_{min} are 0 and 1 in this case [27].

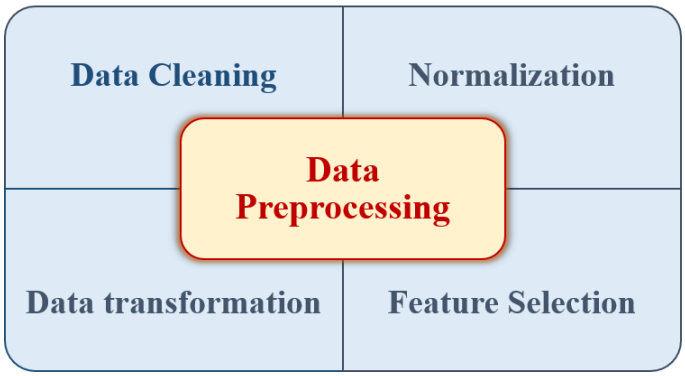

Figure 2.1. Data preprocessing techniques in machine learning.

In data transformation step, data is converted to a useful form to utilize specific machine learning or data mining methods. It is especially required for categorical attributes [26]. One-hot encoding (dummy coding) is the most common encoding system that assign 0 or 1 to represent specific category [28].

Feature selection creates a subset of features from high dimensional data for model building as a preprocessing strategy. A huge amount of data is extensively used to extract knowledge in various fields such as education, healthcare, and social media. Nonetheless, using big data may lead to several issues in machine learning and data mining application. Overfitting can decrease performance of the model on unseen data due to the usage of large number of inputs. Also, it can increase computational cost and there is no need to compute all features if we are able to build our model with low number of inputs. Feature selection is an important step to reduce complexity of the model and get simple and efficient dataset [29]. Alternative feature selection methods were attempted to eliminate redundant features. In this work, trial/error

method, creating a subset of features and random forest feature importance were used as feature selection models.

2.1.2. Performance Metrics

Feature selection, model selection, and preprocessing are gaining a prominent role as research subjects due to the increase in machine learning applications. The comparison of various models is the key issue to interpret viable design recommendations and to elucidate their implications. Performance metrics allow quantifying the applicability and reliability of machine learning models. Although the result of unseen data is substantial, it is important to apply performance measurements to the training and validation set as well. The predictive performance of training data and comparing training and test dataset are required to distinguish potential problems such as overfitting and underfitting. Also, the errors of cross-validation (CV) sets should be consistent. In regression tasks, root mean square error (RMSE) and coefficient of determination(R^2) are widely reported performance metrics [30]. RMSE and R^2 formulas are defined as,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|e_i|^2} \tag{2.3}$$

$$R^2 = 1 - \sum_{i=1}^{n}\frac{|e_i|^2}{|y_{i,true} - \bar{y}|^2} \tag{2.4}$$

where $e_i = y_{i,true} - y_{i,predicted}$, n is the total number of data points and $\bar{y}$ is the mean of y values. In this work, RMSE is used as performance metric since it is sensitive to outliers and gives insight into homogeneity of errors. To observe correlation between actual and predicted values R^2 is determined for each model [30].

2.1.3. Linear Models

Statistical learning or machine learning is a key issue to examine data by using various tools in many fields. Linear regression is a simple and useful approach to predict the quantitative output and interpret the relationship of inputs and output. The linear regression model can be represented as

$$y = \beta_0 + \sum_{j=1}^{n} x_j \beta_j \tag{2.5}$$

where y is dependent variable, x is the vector of inputs, n is number of inputs, β_0 and β_j are coefficients that refers intercept and slope, respectively. The model parameters should be obtained for future predictions. Least squares is commonly used method to pick model coefficients β_0 and β_j. The least squares method chooses the model parameters by minimizing residual sum of squares (RSS), also known as sum of squares error (SSE). The formula of RSS for N data points can be presented as

$$RSS(\beta) = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{n} x_{ij} \beta_j \right)^2 . \tag{2.6}$$

In linear regression, some of the variables can be redundant that leading to unnecessary model complexity. The model interpretability may increase by removing irrelevant variables. Lasso regression ignores these variables by assigning corresponding model coefficients to zero. The lasso estimate can be defined as,

$$\hat{\beta}^{lasso} = \underset{\beta}{argmin} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{n} x_i \beta_i \right)^2 + \lambda \sum_{j=1}^{n} |\beta_j| \right\} \tag{2.7}$$

where (λ) is tuning parameter and greater than zero. The term $|\beta_j|$ is the lasso penalty (also known as l_1 penalty). When λ is large enough, the lasso penalty regulates some of the model coefficients as zero. Hence, the selection of λ is very critical to build an effective model. In general, it is chosen by using cross-validation. The λ that gives the average error is used as optimum λ. In addition, lasso regression can be

performed for variable selection since it sets the unnecessary variable into zero [28,31]. The linear regression models estimate a linear relationship between the dependent and input variables. In some cases, the output variable can be defined by the predictors in a nonlinear relationship. Therefore, alternative machine learning models are discussed.

2.1.4. Regression Tree

Tree-based approaches split the feature space into rectangles, that are subsequently fitted with a simple model (such as constant) for each one. They are straightforward but quite powerful methods. The partitioning method depends on the type of features and splitting procedures such as binary or multi-way splitting. In order to reduce complexity, recursive binary partitioning is used to divide feature space. It is known as greedy approach since it chooses the best split for a particular step. It does not consider the split which leads better results in future steps during the tree-building process. The regression tree algorithm requires deciding on how to split points and variables. The general partitioning process can be seen below,

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R). \tag{2.8}$$

Here, M donates number of partition regions (R_1, R_2,..,R_M) and c_m is constant response. The aim is finding R_m that minimizes the sum of squares (SSE) of f(x). It can be seen that it is computationally infeasible considering each possible partition. Hence, recursive binary partition is performed for all data with selected predictor x_j and split point s to identify pair of half-planes in Equation (2.9),

$$R_1(j,s) = \{x|x_j \leq s\} \qquad\qquad R_2(j,s) = \{x|x_j \geq s\}. \tag{2.9}$$

Then, j and s can be found from Equation (2.10),

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i \in R_1(j,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i \in R_2(j,s)}(y_i - c_2)^2\right]. \tag{2.10}$$

The best pair (j, s) that prone to lower error is found by evaluating all possible variables and split points. These procedure are conducted for all previous regions until the stopping criteria is satisfied.

Furthermore, tree size is an important tuning parameter that affects complexity of model. Large trees may lead to overfitting that causes low prediction accuracy on unseen data while small trees may not distinguish important structure. Hence, the optimal size is needed to achieve satisfactory results. The favored procedure is specifying minimum node size (for instance 5 or 10) to stop splitting.

Regression trees have several advantages such as interpretibility, computing both numerical and categorical predictors and little preporocessing. Also, regression trees are more efficient than linear models to handle nonlinear behaviours. However, they are not effective compared to tree-based ensembles such as random forest and boosting with regard to the accuracy of predictions [28, 31].

2.1.5. Tree-based Ensembles

Tree-based ensembles integrate the predictions of many different trees to give a consensus prediction. Tree-based ensembles offer several advantages in many aspects such as fast computation, high accuracy and little preprocessing. Bagging, random forests, and boosting are useful methods that yield improved accuracy. For the regression task, the final output is the average prediction values that come from individual trees [32, 33]. Bagging is an effective procedure to reduce the variance of statistical learning algorithms. The main drawback of regression trees is high variance, the results may change substantially according to splitting points. Bagging method creates different samples from the training set and evaluates average results of these sets to handle high variance issues. Averaging is beneficial treatment for tree-based models since they are noisy [28, 31].

Random Forest is promising approach that offers advanced method of bagging by decorrelating the trees. The aim of random forests is to enhance variance decreasing of bagging method by reducing correlation of tree to obtain more reliable model. In bagging method, the variance reduction may arise from the similarity of trees. For instance, all trees use the same important predictor as a splitting variable that means bias of the bagged tree is the same with a single tree. Random forest selects random input variables in the tree-growing step to decorrelate trees. The subset size of random variables is an important parameter that affects the performance of the model. For regression, one-third of the total variables is the default value for number of selected random variables. After B trees are grown as a random forest, predictions of regression tree can be defined as

$$\hat{f}_{rf}^{B} = \frac{1}{B}\sum_{b=1}^{B} T(x;\theta_b) \tag{2.11}$$

where B is the number of trees and θ_b denotes the bth random forest tree in terms of split points and variables [28, 31]. Out-of-bag (OOB) error estimation is a prominent characteristic of the random forest approach. Random forest uses one-third of the instances as out-of-bag samples to measure performance of the model. Hence, there is no need to leave out any sample as a validation set [34]. These out-of-bag samples give predictions for each instance. In the regression task, the final prediction is attained by using the average of these predicted values. Also, OOB samples can be used to determine importance of variables. The importance of each predictor can be obtained using the residual sum of squares. The RSS is recorded to evaluate splits over a particular variable by using average of all trees. A large decrease in RSS implies importance of predictor [28].

Boosting is a common method for machine learning algorithms for regression task. It is successfully integrated into regression trees to mitigate some problems. Boosting operates the same principle as bagging. Bagging method proceeds independently of other trees whereas boosting utilizes sequential growing by using the information of previous trees. The main limitation of single trees is low accuracy due to the piecewise

constant approximation. Gradient boosting of regression trees posses high accuracy, competitiveness and robustness as an ensemble model [28, 35].

Gradient boosting seeks the new tree that gives better results compared with the previous trees. The new model is constructed by using the residuals of the current model. The residuals are determined by using a specific loss function which is widely least squares for regression. Briefly, the new tree focus on where the algorithm does not fit well. The shrinkage parameter slows the learning rate of the model to screen residuals in detail. The boosting model learns slowly but shows high performance. The output value of the model can be represented as

$$\hat{f}_{gb}^{B} = \frac{1}{B}\sum_{b=1}^{B} \lambda T(x;\theta_b) \tag{2.12}$$

where λ is shrinkage parameter [28, 33].

Gradient boosting employs only small trees as weak learners different from the random forest model. The size of individual trees is controlled by interaction depth which is a tuning parameter of gradient boosting. Also, the number of trees should be examined as a tuning parameter by using cross-validation since the high number of trees may lead to overfitting. Another tuning parameter is shrinkage (λ) that defines the learning rate of boosting. The selection of λ affects the number of trees, very small λ corresponds to large number of tree to achieve desired performance [28, 33].

2.1.6. Artificial Neural Networks

Artificial neural network with a computational structure that consists of several highly interconnected nodes that gives a dynamic response as an output [36]. Neural networks have received attention due to the ability of learning non-linear and complex problems and flexibility of application procedure for large datasets [37]. Neural networks (also known as feedforward neural networks or multi-layer perceptron) comprise an input layer, hidden layers, and output layer. This model is called feedforward since

the information is conducted input layer to output. There is no feedback from the output of the model into itself. The final layer of the model is called output layer. The behavior of other layers is determined by algorithm to obtain desired output. These layers are known as hidden layers [38]. Mathematical modelling of artificial neural network can be represented as

$$a = \sum_{i=0}^{D} w_i \mathbf{x}_i + w_0, \tag{2.13}$$

where x_i is input value for i goes from 1 to D (number of inputs), w_i is weight and w_0 is bias. The output can be obtained by using nonlinear activation function that is shown in Equation (2.14):

$$y = f(a). \tag{2.14}$$

Here, f is activation function and y is output value [38, 39]. Feedforward networks include hidden layers that requires selection of activation function to determine hidden layer values. The selection of activation function depends on nature of data and target variable. Sigmoid and hyperbolic tangent are common choices as an activation function. Also, some design parameters are needed to train feedforward neural networks such as objective function (also called error or cost function), number of hidden layers , number of neurons and choosing the optimizer. Common error functions are sum of squared errors (SSE) and mean square error (MSE). The goal is finding the system that gives lower error for the training set [39, 40].

The training process of neural networks involves updating the weights to obtain desired input/output relationship. The aim is to train the network using given data to minimize objective function. The back-propagation method is widely used as a training algorithm for neural networks. Nevertheless, there are several optimization methods that may achieve better results to train the neural network such as steepest descent, Levenberg–Marquardt method, and quasi-Newton method. In general, Levenberg–Marquardt algorithm is used in order to train artificial neural networks [40, 41].

The Levenberg–Marquardt algorithm was independently developed by Kenneth Levenberg and Donald Marquardt. In the artificial neural-networks field, this algorithm which provides a solution to the problem of minimizing a nonlinear function is suitable to train small and medium-sized problems. There are a lot of methods for neural network training. The steepest descent algorithm known as error backpropagation (EBP) algorithm was important breakthroughs for training neural networks. However, it is inefficient method since it converges slowly because of the oscillation when approaches minima. The steepest descent method is improved by the Gauss–Newton algorithm which provides fast convergence [42].

The Levenberg–Marquardt algorithm combine the steepest descent and the Gauss Newton algorithm which provides fast convergence and stability. The working principle of the Levenberg–Marquardt method is that the algorithm performance steepest descent method around the area with complex curvature, and it switches to the Gauss-Newton algorithm near the local curvature. The Levenberg–Marquardt algorithm can be presented as

$$\mathbf{H} = (\mathbf{J_k}^T \mathbf{J_k} + \mu \mathbf{I}) \tag{2.15}$$

$$\mathbf{w_{k+1}} = \mathbf{w_k} - (\mathbf{J_k}^T \mathbf{J_k} + \mu \mathbf{I})^{-1} \mathbf{J_k} \mathbf{e_k}, \tag{2.16}$$

where $\mathbf{H}$ is the Hessian matrix, $\mathbf{J}$ is the Jacobian matrix, $\mathbf{I}$ is identity matrix, μ is combination coefficient and always positive and $\mathbf{e}$ is the error. When combination coefficient μ is large, the algorithm approaches the steepest descent, when μ is small the algorithm approximates to the Gauss-Newton algorithms [42]. Specifications for various algorithms are summarized in Table 2.1. Here, a refers step size and $\mathbf{g}$ is first-order derivative of error function (gradient).

2.2. Genetic Algorithm

Evolutionary algorithms (EAs) are heuristic population-based models that employ biology-inspired techniques including natural selection, crossover, and mutation.

Table 2.1. Specifications for different algorithms [42].

Algorithms	**Update Rules**	**Convergence**	**Computation Complexity**
EBP	$\mathbf{w_{k+1}} = \mathbf{w_k} - a\mathbf{g_k}$	Stable,slow	Gradient
Newton	$\mathbf{w_{k+1}} = \mathbf{w_k} - \mathbf{H_k}^{-1}\mathbf{g_k}$	Unstable,fast	Gradient&Hessian
Gauss-Newton	$\mathbf{w_{k+1}} = \mathbf{w_k} - (\mathbf{J_k}^{\mathbf{T}}\mathbf{J_k})^{-1}\mathbf{g_k}$	Unstable,fast	Jacobian
Levenberg-Marquardt	$\mathbf{w_{k+1}} = \mathbf{w_k} - (\mathbf{J_k}^{\mathbf{T}}\mathbf{J_k} + \mu\mathbf{I})^{-1}\mathbf{g_k}$	Stable,fast	Jacobian

They are based on Darwin's evolutionary theory that is survival of the fittest. Heuristic optimization algorithms are utilized to identify the next solution candidate or next individual to be produced by using previous information collected by the algorithm [43]. In the 1960s, genetic algorithm (GA) was devised and developed by John Holland. It was a breakthrough in terms of the population-based method with selection, crossover, and mutation operator. The approach of searching among a set of candidate solutions for desired outputs is called search space and it is widely used in computer science. Genetic algorithms may utilize to increase search space by using reproduction methods. Genetic algorithms consist of basic implementations such as population, selection procedure according to the fitness function, reproduction procedures including random mutation and crossover to create new offspring. In genetic algorithm, chromosome denotes candidate solutions of the problem, and crossover is used for exchanging some sub-part of two chromosomes. The GA applies a fitness function that specifies a score or fitness for each chromosome in the present population. The score of a chromosome is defined by how well the chromosome handles the problem. Genetic algorithm tends to select fitter chromosomes for the reproduction process. The genes refer to single bits or short blocks that encode the candidate solution. The mutation changes some arbitrary genes to create new offspring [44].

A simple genetic algorithm is shown in Figure 2.2. Each iteration that is depicted in Figure 2.2 is called generation. Moreover, there are additional procedures to improve the performance of genetic algorithms. Elitism was introduced by Kenneth De Jong in 1975 which is an additional selection method. Elitism allows for keeping a certain number of the best individuals in each generation. It is a beneficial process since such

individuals may decompose by crossover and mutation operators [44].

A simple genetic algorithm

1. Begin with randomly generated population that consist of n individuals

2. Calculate the fitness function for each chromosome in the current population

3. Operate the following steps until n children solution have obtained

(i) Select two parent chromosomes from the population based on fitness function. The same chromosomes can be chosen more than once as a parent

(ii) Crossover the selected pair at a random point for specified crossover rate (p_c), also known as crossover probability to form new offspring

(iii) Mutate the two offspring for a given mutation probability (p_m)

4. Insert new candidate solutions to form new population

5. Go to step 2

Figure 2.2. Representation of simple genetic algorithm.

Selection is the basic inspiration of genetic algorithm that improves poor solutions. The fittest individuals have a better chance of achieving food and mating in nature. Hence, they inherently transfer their genes to the next generation. Genetic algorithm similarly select parent individuals by using their fitness value. Tournament selection and roulette wheel selection are common methods. Roulette wheel selection assigns a probability for each individual according to fitness value and the higher probability has a higher chance to contribute next generation. Also, poor individuals may move to the next generation coincidentally. However, such individuals should be kept for the sake of diversity [44]. Tournament selection is another important selection mechanism. Tournament selection organizes several tournaments among competitors, where s is the tournament size. The winner is the individual that has a higher fitness value in a certain tournament. The winners of tournaments are selected as parent solutions to form new offspring. Hence, the selected solutions as parents provide a higher average fitness value than population which increases the performance of the genetic algorithm [45].

Crossover is the main operator of genetic algorithm to obtain children solutions. There are several techniques, yet, the common ones are single and double point crossover. In the single point crossover, the sub-part of parent chromosomes are changed according to crossover point. In the double-point crossover utilize two crossover point and between these points in chromosomes are changed. Crossover and mutation operators are presented in Figure 2.3 and Figure 2.4, respectively.

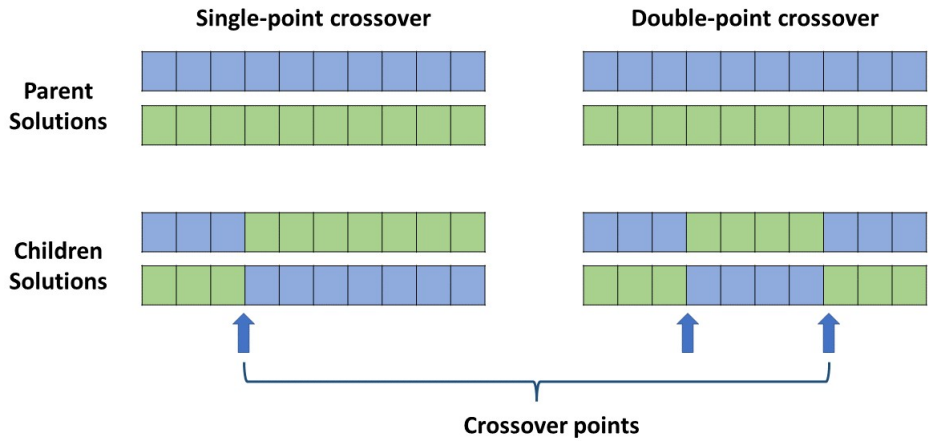

Figure 2.3. Representation of single and double crossover procedures.

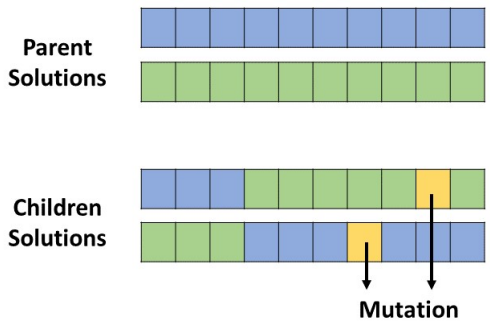

Figure 2.4. Mutation operator.

GAs are prominent methods for handling difficult technological problems. Many real-world problems involve nonlinear relationships. Analytical or optimization techniques may be limited for nonlinear problems due to the computational cost and instability. For instance, using matrix inversion or partial derivatives can be difficult and time-consuming. Therefore, genetic algorithm may be useful to eliminate such

problems. Genetic algorithms explore search space by using reproduction methods with the information gain. It is a promising and low-cost model to obtain an optimum solution [46]. Furthermore, the method the variables are coded is critical to the effectiveness of the genetic algorithm. A real coded genetic algorithm (RCGAs) using real numbers converges to the optimum faster than a binary-coded genetic algorithm. Genetic algorithm may handle both binary and real coded problems. Therefore, it is an efficient algorithm to solve mixed-integer optimization problems [47].

2.3. Artificial Neural Networks-Genetic Algorithm Algorithm Model

Artificial intelligence algorithms are computational methods that can perform complex and intelligent mechanisms of the human brain. Artificial neural networks are brain-inspired algorithm that mimics the working principle of the brain. ANN is a powerful method to discover nonlinear relationships among input and output features. However, there are some challenges to designing and training neural networks. Learning algorithm parameters and network structure can affect the performance of neural networks. Genetic algorithm is a search algorithm and inspired by basic genetic information and natural selection such as mutation, crossover, and selection. Integration of ANN and GA provides a more accurate model and optimization procedure to achieve satisfactory results. ANN-GA hybrid model is applied for several fields such as biological processes and optimization of chemical processes [37, 48–50].

Cook *et al.* reported the NN-GA model as an optimization for process parameter optimization. The output of neural networks is used as fitness function of genetic algorithm. It is obtained that the NN-GA hybrid system was successfully identified process parameters under various conditions and at stages of the process to achieve a desired internal bond [49]. Salehi *et al.* proposed that integrated MLP-GA algorithm maximize biosynthesis of paclitaxel through several inputs such as concentration and harvesting time. It is found that the developed MLP-GA model determines the input variables that yield maximum paclitaxel biosynthesis [50]. The reported results appear promising and ANN-GA hybrid model may be utilized as an effective optimization

tool. The schematic of Artificial Neural Network-Genetic Algorithm (ANN-GA) hybrid model is represented in Figure 2.5.

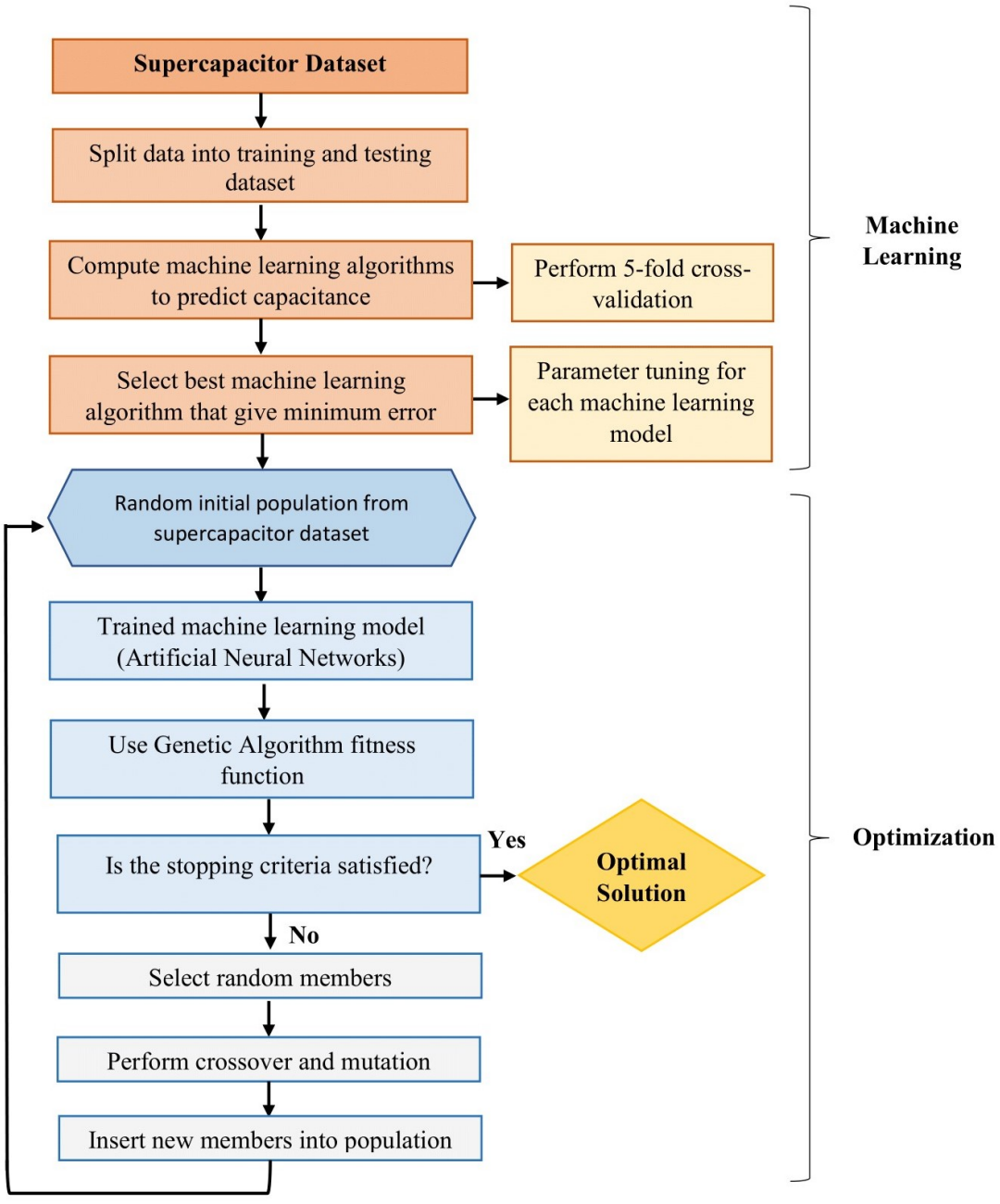

Figure 2.5. Schematic of ANN-GA hybrid system.

Figure 2.5 indicates brief outline of the optimization procedure. Firstly, dataset is divided into training and testing set to measure the performance of machine learning algorithms. Different machine learning algorithms are computed with tuned parameters to predict capacitance by using input variables. The high-performance algorithm is selected as a predictive model according to specified performance metrics. Random initial population is formed that encompasses the lower and upper limits of each variable. The output of trained artificial neural networks is utilized as fitness function of genetic algorithm. Binary tournament selection is implemented to choose parent solutions concerning fitness values. Crossover and mutation operators are performed to obtain children solutions and these new offspring are inserted into the population. These sequential steps are conducted until the stopping criteria are reached which is maximum number of generations. Also, it is inevitable to define boundaries of search space to avoid a solution impossible to implement. Hence, constraints of input variables are assigned to the genetic algorithm. These constraints are identified based on expert knowledge, literature information, and characteristics of variables. In order to specify the relationship of two inputs, the big-M method is utilized.

2.3.1. Mixed-Integer Nonlinear Programming

Optimization has become an advantageous tool in various domains such as scheduling/planning, design problems, and process system engineering. It has emerged as an academic study and provided significant contributions to the industry in terms of economics and engineering. Optimization problems can be classified through continuous and discrete variables. Also, there are two important parameters related to optimization problems that are convexity and differentiability. Nonconvex problems may lead to local optima. The main motivation of the optimization procedure is finding the global optimum. Global optimization especially has focused on design problems since nonconvexities in the problem may result in suboptimal solutions due to the flexible constraints of several variables. Hence, robust methods are required to solve all kinds of optimization problems [51]. Mixed-integer nonlinear programming is a substantial method to handle optimization problems that consist of continuous and discrete vari-

ables. The general form of mixed-integer nonlinear problems (MINLP) is given as,

$$min\ Z = f(x,y)\ subject\ to \begin{cases} h(x,y) = 0 \\ g(x,y) \leq 0 \\ x \in X,\ y \in \{0,1\}^m. \end{cases} \tag{2.17}$$

Here, $f(x,y)$ is objective function , $h(x,y) = 0$ linear or nonlinear equality constraints, $g(x,y) \leq 0$ linear or nonlinear inequality equations that specify constraints of variables, x refers to continuous variables, and y are integer (discrete) variables that take 0 or 1. For instance, the discrete variable denotes the selection of heat exchangers in process system engineering or material selection in design problems [52].

Many scientific applications and industrial operations involve a combination of continuous and discrete decision variables with the nonlinear objective and/or constraint functions. MINLP problems deal with nonconvex functions, discrete variables, and nonlinear functions to find the optimum solution. There are several methods including branch-and-bound, outer approximation, and disjunctive programming to solve MINLP problems. For convex MINLP problems, these algorithms may result in promising solutions. However, nonconvex MINLP problems create additional problems since the nonconvexity leads to local optima. Therefore, heuristic methods may be utilized to obtain a feasible solution rapidly. Genetic algorithm is an alternative approach to solve MINLP problems thanks to the effective search and improving objective function generation by generation [53–55].

Mixed-integer programming is commonly utilized with model indicators for identifying logical constraint is active. The big-M representation is a simple and widely used method to deal with such constraints. The most straightforward approach is big-M formula that involves multiplying the binary indicator variable by a very big constant M to activate or deactivate constraints of the problem [56, 57].

The mathematical formulation can be represented as

$$g_i(x) \leq M_i(1 - b_i) \tag{2.18}$$

where, $g_i(x)$ represents inequality constraint, M_i is very large positive number and b_i refers binary indicator {0,1}. It can be seen that if $b_i = 1$, then $g_i(x) \leq 0$ is forced. Otherwise, if $b_i = 0$, constraint is satisfied for any value that is feasible for x [57].

The essential part is the selection of M value. M should be big enough to outweigh the other values in the problem. However, very large numbers increase the computational cost. In general, a single M is used for each variable, yet, it depends on the nature of the problem. In some cases, more than one M constant is required for variables. Also, it is commonly used to handle constraints in optimization software such as IBM Cplex and Gurobi [57, 58]. In this work, modified big-M method is utilized to handle linear inequality constraints that are given in Figure 2.6.

Modified big-M method

If $Z \leq N$, **then** $Y \leq A$;

1. $Z - N \leq Mb$

2. $Z - N \leq M(1 - b)$

3. $Z - N \geq -Mb$,

where Z and Y are variables, N and A is numerical values, M is a large positive constant and b is binary variable.

Figure 2.6. Modified big-M formulation for linear inequality constraints.

In this thesis, a supercapacitor design that yields optimum capacitance is proposed by using hybrid ANN-GA. Supercapacitor design parameters include the se-

lection of main components/related characteristics and operating conditions. These features consist of both discrete and continuous variables. Categorical attributes such as electrode material and current collector can be given as an example for discrete variables. For instance, potential window, specific surface area of the electrode, and electrolyte concentration are continuous variables. Also, the objective function is non-linear since artificial neural networks is used as the machine learning algorithm. In the light of this information, the proposed method may be inferred as mixed-integer nonlinear problem. The objective function and several constraints are utilized to construct the model that gives the feasible and optimum solution. The side constraints are given to specify the lower and upper limits of all variables. Linear equality constraints are used to impose the selection of a single material or method for each categorical feature. Also, linear inequality constraints are employed to determine the relationships of two features that enable real-world implementation. These inequality constraints are formulated by introducing a big number M and artificial binary variables for each constraint. The problem formulation of the hybrid ANN-GA model is given as follows,

Objective Function:

$$f(x) = \mathbf{max}\ Capacitance \tag{2.19}$$

subject to

Linear equality constraints

$$\sum_{i=11}^{24} x_i = 1\ for\ Electrode\ Material \tag{2.20}$$

$$\sum_{i=25}^{32} x_i = 1\ for\ Preparation\ Method \tag{2.21}$$

$$\sum_{i=33}^{39} x_i = 1\ for\ Separator \tag{2.22}$$

$$\sum_{i=40}^{50} x_i = 1\ for\ Current\ Collector \tag{2.23}$$

Linear inequality constraints

$$If\ HeatT \leq 0.5,\ then\ IDIG \leq 0.41 \tag{2.24}$$

$$If\ SSA \leq 0.71,\ then\ Oconcentration \geq 0.52 \tag{2.25}$$

$$If\ Electrolyte\ Concentration \geq 0.5,\ then\ APW \geq 0.25 \tag{2.26}$$

$$If\ Electrolyte\ Concentration \leq 0.32,\ then\ APW \geq 0.58 \tag{2.27}$$

$$If\ Nconcentration \geq 0.24,\ then\ IDIG \leq 0.42 \tag{2.28}$$

$$If\ Pore\ Volume \geq 0.87,\ then\ SSA \leq 0.28 \tag{2.29}$$

$$If\ Oconcentration \geq 0.42,\ then\ IDIG \leq 0.37 \tag{2.30}$$

$$If\ Pore\ Volume \geq 0.62,\ then\ IDIG \leq 0.32 \tag{2.31}$$

$$If\ Oconcentration \geq 0.34,\ then\ Nconcentration \leq 0.29 \tag{2.32}$$

$$If\ Pore\ Volume \geq 0.35,\ then\ Oconcentration \leq 0.45 \tag{2.33}$$

Side constraints

$$0 \leq x_1 \leq 1\ for\ HeatT \tag{2.34}$$

$$0 \leq x_2 \leq 1\ for\ SSA \tag{2.35}$$

$$0.10 \leq x_3 \leq 1\ for\ IDIG \tag{2.36}$$

$$0 \leq x_4 \leq 0.31\ for\ Nconcentration \tag{2.37}$$

$$0 \leq x_5 \leq 1\ for\ Oconcentration \tag{2.38}$$

$$0.01 \leq x_6 \leq 0.65\ for\ Pore\ Volume \tag{2.39}$$

$$0.06 \leq x_7 \leq 0.8\ for\ Electrolyte\ Concentration \tag{2.40}$$

$$0.5 \leq x_8 \leq 0.5\ for\ Normalized\ Scan\ Rate \tag{2.41}$$

$$0.17 \leq x_9 \leq 1\ for\ Absolute\ Potential\ Window \tag{2.42}$$

$$0 \leq x_{10} \leq 1\ for\ Binder\ Concentration \tag{2.43}$$

$$0 \leq x_j \leq 1\ for\ categorical\ and\ binary\ variables\ (j = 11...60) \tag{2.44}$$

where x is the decision variables of the genetic algorithm, M is a large number and b is

binary variable of the big-M method. Also, except that continuous features (x_1 *to* x_{10}), the others are integer variables. Artificial neural networks is computed by using deep learning toolbox and genetic algorithm optimization is proceeded by using the "ga" function of MATLAB.

3. RESULTS AND DISCUSSION

In this section, dataset information and preprocessing methods are given to prepare dataset for further procedures. Outlier detection, normalization, cross-validation, and feature selection are performed as preprocessing. Several feature selection methods are applied such as random forest predictor importance, trial/error method and creating subset of features. Different machine learning algorithms are selected to compute regression task by using given dataset. The aim is to compare the performance of linear regression, penalized regression approach (lasso), tree-based ensembles and artificial neural networks to predict capacitance and obtain optimum capacitance value. The performance of the algorithms is measured for specified parameters and results are compared. Root mean square error (RMSE) used as performance metric to evaluate results. Also, the outputs of proposed hybrid Artificial Neural Network/Genetic algorithm-based optimizer are represented. All computational procedures are carried out by using MATLAB software.

3.1. Dataset Information

Domain knowledge is important parameter to assess results of machine learning algorithms. It is also useful for outlier detection and to understand impossible data combinations. For example, potential window of aqueous electrolytes cannot be 6 V since the electrolysis of water occurs at 1 V. If there is a such problem in the dataset, these instances can be discarded. Therefore, the characteristics and relationships of features should be carefully examined.

In this work, the dataset is used that includes 2189 data points collected from 110 published articles for supercapacitors [59]. The dataset includes both categorical and numerical features that described supercapacitors such as electrode material, preparation method and specific surface area. Capacitance is the target attribute of the dataset and machine learning algorithms are computed to predict capacitance by

using 22 predictors. These predictors related with electrode/electrolyte materials and operation conditions. Heat treatment, specific surface area, ID/IG ratio, pore volume, oxygen/ nitrogen/boron/phosphorous concentration are related with electrode materials. Electrolyte concentration, solvent anion/cation volume, solvent volume and solvent dipole moment related with electrolyte material. Binder concentration, separator and current collector are independent features. Scan rate and potential window are operational conditions. However, potential window and electrolyte type are highly correlated since the voltage affects the electrolyte degradation according to type of electrolyte. Whereas higher voltages are suitable for ionic liquids, aqueous electrolytes cannot be operated at high voltage. Features that are described for supercapacitors are summarized in Table 3.1. Also, categorical variables and statistical information of numerical features are listed in Table 3.2 and Table 3.3, respectively.

Table 3.1. Dataset Information.

SUPERCAPACITOR DATASET	**INFORMATION**
Capacitance (Target)	Energy storage ability of capacitor (F/g)
ElectMG	Electrode material of supercapacitor Ex: Carbon materials, metal oxides or polymers
HeatT	Activation temperature of electrode material (°C)
SSA	Specific surface area of electrode material (m^2/g)
IDIG	Intensity ratio of D and G gaps
Nconcentration	Nitrogen amount of electrode material (%)
Oconcentration	Oxygen amount of electrode material (%)
Pconcentration	Phosphorus amount of electrode material (%)
Bconcentration	Boron amount of electrode material (%)
Pore Volume	Pore volume of electrode material (cm^3/g)
PrepMG	Preparation method of electrode material Ex: Thermal, microwave exfoliation, hydrothermal
Electrolyte Types	Electrolyte includes solvent and positive/negative ions
Electrolyte Concentration	Concentration of electrolyte material (M)
Normalized Scan Rate	Scan rate of the process (mV/s)
Abs Potential Window	Operation voltage range(V)
Separator	Separator is soaked in electrolyte and separates positive and negative electrodes Ex: Celgard, Glass fiber, Cellulose
Current Collector	It collects electrons from electrode material Ex: Ni,Al,Pt,Fe
Salt Anion Volume	Anion volume of salt in the electrolyte Ex: 18.01 for KOH
Salt Cation Volume	Cation volume of salt in the electrolyte Ex: 40.59 for KOH
Solvent Dipole Moment	Dipole moment for each solvent Ex: 1.85 for water, 3.92 for acetonitrile
Solvent Volume	Solvent volume of electrolyte
Binder Concentration	Binder is added between electrode and current collector (%) Ex: PTFE, PVDF, PVDC, PP

Table 3.2. Categories of features.

Categorical Features	**Number of Categories**	**Categories**
ElectMG	14	Activated carbon (AC) Carbide derived carbon (CDC) Carbon nanotubes (CNTs) Carbon nanofiber Carbon nanospheres (CNs) CNs/CNTs CNs/Graphene Carbon xerogel Graphene Graphene/AC Graphene/CNTs Graphite Metal organic framework Mxenes
PrepMG	8	Arc-discharge Chemical activation Chemical vapor deposition (CVD) Chlorination Hydrothermal Microwave exfoliation Thermal activation Thermal and chemical activation
Electrolyte Type	3	Aqueous Organic Ionic liquids

Table 3.2. Categories of features (cont.)

Categorical Features	**Number of Categories**	**Categories**
Separator	7	Aluminum oxide (Al_2O_3) Celgard Cellulose Glass fiber No separator PTFE Polyproylene
Current Collector	11	Silver (Ag) Aluminum (Al) Iron (Fe) Nickel (Ni) Platinum (Pt) Titanium (Ti) Stainless steel Al/Fe Au/Pt Au/Stainless steel Carbon substrates

3.2. Data Preprocessing

Multiple preprocessing methods are attempted to obtain more robust results. Outlier detection, normalization, cross-validation and feature selection are performed to eliminate irrelevant data. There are 2189 instances and 22 feature that are extracted from experimental published paper. Capacitance values in the dataset are examined since the capacitance is target feature. It is observed that capacitance predictions for larger than 300 did not give satisfactory results for machine learning models.

Table 3.3. Statistical information of numerical features.

Numerical Features	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**
Capacitance (Target)	0.08	295	82.44	60.34
HeatT	25	1000	672.77	250.70
SSA	0.99	3523	1455	934.76
IDIG	0.35	4	1.27	0.64
Nconcentration	0	20.55	2.24	4.35
Oconcentration	0	28.81	8.26	4.01
Pconcentration	0	1.90	0.01	0.12
Bconcentration	0	12.90	0.07	0.94
Pore volume	0.003	4	1.174	0.63
Electrolyte concentration	0.5	8.1	4.16	2.12
Normalized scan rate	0	1	0.53	0.37
Abs Potential Window	0.4	4.8	1.81	1.16
Salt Anion Volume	18.01	261.27	73.25	85.37
Salt Cation Volume	9.03	185.55	79.14	59.08
Solvent Dipole Moment	1.85	12.26	2.69	1.25
Solvent Volume	19.61	100.2	32.36	17.96
Binder Concentration	0	15	4.49	4.27

The number of data points larger than 300 is 14. Therefore, these data points are discarded since supercapacitor dataset includes 2189 instances and there is no sufficient data to learn model and make predictions for high values of capacitance. Outlier detection, also known as anomaly detection, is used to find unexpected cases or exceptional data points that different from the majority patterns [26]. Therefore, outlier detection is performed to eliminate outlier data points and build robust model. 168 outlier data points are discarded from dataset for HeatT, Oconcentration, Pore Volume and Normalized Scan Rate (NSR) in total and 2021 instances are used to build machine learning models.

In general, features that have larger range may affect data analysis process and may result in greater impact on results. Normalization should be used to prevent such a problem. Normalized data is used to obtain minimum bias and get rid of scale inconsistencies. Normalization is applied in the range of [0, 1] for numerical variables according to minimum and maximum values of features. Inputs are normalized to obtain more interpretable train and test errors for machine learning and ANN-GA algorithms. Capacitance (output) is used without normalized. Genetic algorithm results are obtained with normalized values. Hence, parameters that give maximum capacitance values converted into real values by using denormalization. Thus, we can control if parameters are in admissible range or not.

The dataset includes important categorical features for optimum supercapacitor design such as electrode material, preparation method of electrode, separator, current collector and electrolyte type. Categorical variables should be transformed into numerical form since some machine learning algorithms such as linear regression and neural networks cannot handle categorical variable. Therefore, one-hot encoding is implemented to deal with categorical variables. One-hot encoding representation of electrolyte type that contains three categories can be seen in Figure 3.1.

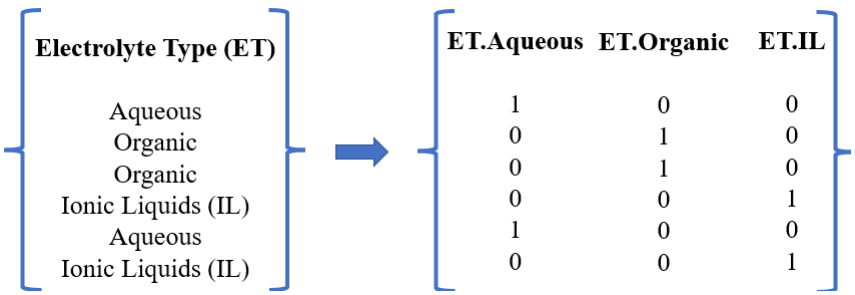

Figure 3.1. Representation of one-hot encoding.

Cross-validation is widely used method to assess accuracy of train and test predictions. Ideally, the validation set can be used to evaluate performance of the model due to the size of the data. However, the datasets are often scarce and it is not possible put aside such a validation set. In order to solve this problem, k-fold cross-validation

is attempted and k-1 part is used to build model and kth part is used to test performance of the model. In general, five- or ten-fold cross-validation are recommended [60]. 5-fold cross-validation for linear/lasso regression, regression tree, random forest, gradient boosting and 10-fold cross-validation for artificial neural networks are performed by manually. Also, cross-validation error deviations are determined for each algorithm to check accuracy of train/test predictions. 5-fold cross validation is done by manually for Linear/Lasso Regression, Regression Tree, Random Forest, and Gradient Boosting models. Train data (1617 instances) is divided 5 equal parts by using randomly selected instances. Each part is used for testing and the rest (4 part) for training set. The results are consistent since the error deviations are not significant that can be seen in Table 3.4. CV1 is the best regularized model.

Table 3.4. RMSE errors for 5-fold cross-validation sets of machine learning algorithms (80 % training set).

Train	**Linear Reg**		**Lasso Reg**		**Reg Tree**		**RF**		**GB**	
CV	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
CV1	39.478	38.800	40.858	39.281	21.425	29.797	20.275	26.185	23.323	27.150
CV2	38.660	42.178	39.392	43.347	22.778	31.380	20.337	28.692	22.802	31.341
CV3	39.452	39.275	40.595	39.478	20.644	31.258	19.269	25.428	23.206	27.920
CV4	39.363	39.344	40.554	40.467	21.416	32.450	18.680	30.879	22.043	32.546
CV5	38.583	42.811	40.107	43.560	20.991	31.099	20.252	25.182	23.030	26.818
Average	39.098	40.482	40.301	41.227	21.450	31.196	20.185	27.478	22.881	29.155
Variation	0.211	3.470	0.331	4.340	0.656	0.894	0.342	6.383	0.250	6.694
SD	0.460	1.863	0.576	2.083	0.810	0.946	0.500	2.587	0.508	2.613

Furthermore, 10-fold cross-validation is computed to evaluate artificial neural networks. Also, sigmoid and hyperbolic tangent are used as activation function. 10-fold cross-validation results can be seen in Table 3.5 and Table 3.6 for hyperbolic tangent and sigmoid activation function, respectively.

Table 3.5. 10-fold cross-validation for artificial neural networks according to number of neurons with hyperbolic tangent activation function.

Neuron	**Average RMSE**		**Variance**		**SD**	
Number	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
10	26.899	30.981	65.955	46.958	7.884	6.704
11	27.550	31.400	65.475	50.335	7.858	6.841
12	26.499	31.004	74.799	56.885	8.230	7.157
13	25.126	29.919	76.531	62.872	8.559	7.675
14	23.827	28.635	65.606	49.016	7.902	6.802
15	23.222	23.554	74.039	47.983	8.381	6.782
16	24.219	28.975	86.960	59.309	8.965	7.474
17	23.672	28.975	102.856	78.130	9.656	8.613
18	24.920	30.029	705.445	69.724	10.083	8.198
19	23.392	28.604	99.739	65.427	9.593	7.625
20	22.173	27.811	87.552	58.830	8.908	7.264

3.2.1. Feature Selection Methods

3.2.1.1. Trial/Error Method. Feature selection decrease computational cost and storage while preventing information loss by removing irrelevant and useless features in real world data [29]. Trial/error method is performed to evaluate individual effects of features and detect unnecessary input variables. The criterion that defines feature importance is the root mean square error of the testing set. Features are discarded one by one and the change in test error is observed. If removing a specific feature is reduced the error, the variable is eliminated from the dataset. If not, the feature is added into data and the process is repeated with different features iteratively [61]. In machine learning algorithms, %80 training ratio and normalized data are used. In this case, if the test errors increased in almost all machine learning algorithms when the feature is discarded, the feature may be considered substantial to predict capacitance. The results are represented in Table 3.7.

Table 3.6. 10-fold cross-validation for artificial neural networks according to number of neurons with sigmoid activation function.

Neuron	**Average RMSE**		**Variance**		**SD**	
Number	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
10	25.221	29.588	52.920	38.717	7.190	6.110
11	25.301	29.639	66.022	54.544	7.889	7.134
12	23.916	29.245	59.138	41.902	7.473	6.044
13	23.979	28.698	61.405	50.046	7.435	6.782
14	22.591	28.333	57.443	46.879	7.283	6.490
15	22.977	28.173	78.144	53.670	8.563	7.061
16	23.108	28.394	71.099	54.526	8.040	7.007
17	22.127	27.507	52.554	41.320	6.967	5.956
18	21.525	27.167	68.207	42.913	7.894	6.304
19	20.380	26.053	71.895	56.597	7.852	7.055
20	20.653	26.538	59.905	39.192	6.940	5.645

The results give a rough idea about importance of predictors. ElectMG, SSA, Nconcentration, PrepMG, Absolute Potential Window (APW), Separator, Current Collector and Salt Anion Volume are increased the test errors of machine learning algorithms that can be observed from Table 3.7. Hence, these features can be considered as important to predict capacitance. However, this method is weak procedure, because when we combine or discard two or three feature together, they can create major impact on dataset. To eliminate such possibility, random subset features are generated to evaluate relationship of different features.

3.2.1.2. Creating Random Subset Features. Random subset feature sets are created to evaluate relationship of different features. Regression tree is used as machine learning algorithm since it does not require one-hot encoding and can handle categorical features. 1000 scenarios are generated for randomly selected subset features.

Table 3.7. Trial/error method for feature selection by using test root mean square errors.

SC DATASET	**ERRORS**					
	Linear	**Lasso**	**RT**	**RF**	**GB**	**ANN**
All data	39.333	39.474	31.806	25.753	25.949	25.020
-ElectMG	41.354	41.792	31.074	25.991	25.673	25.865
-HeatT	39.510	39.836	30.789	25.917	26.965	25.553
-SSA	40.166	40.290	30.133	27.201	26.559	27.440
-IDIG	39.302	39.474	31.806	25.864	25.949	38.910
-Nconcentration	39.420	39.576	32.138	26.120	26.158	49.213
-Oconcentration	39.194	39.440	31.417	25.930	26.289	19.799
-Pconcentration	39.327	39.447	31.761	25.777	25.824	26.134
-Bconcentration	39.507	39.660	31.806	25.704	25.949	18.384
-Pore Volume	39.291	39.474	31.806	25.818	25.956	18.868
-PrepMG	40.172	41.325	30.703	25.767	25.926	30.013
-Electrolyte Type	39.627	39.655	31.511	25.876	26.541	30.643
-Electrolyte Concentration	39.146	39.131	33.571	25.856	26.344	20.347
Norm Scan Rate	39.521	39.652	33.580	26.060	26.352	20.010
-Abs Potential Window	41.073	41.364	32.826	30.245	28.665	25.159
-Separator	44.597	44.798	37.189	28.255	27.524	24.413
-Current Collector	39.383	39.981	33.604	25.577	26.014	29.240
-Salt Anion Volume	39.327	39.634	32.883	26.785	27.166	40.175
-Salt Cation Volume	39.430	39.975	30.250	26.290	26.244	30.610
-Solvent Dipole Moment	39.570	39.526	31.277	25.935	26.177	26.058
-Solvent Volume	39.439	39.728	31.806	25.874	26.126	44.609
-Binder Concentration	40.022	39.974	31.367	25.951	26.524	28.548

Average and minimum train and test errors with respect to number of inputs are demonstrated Figure 3.2.

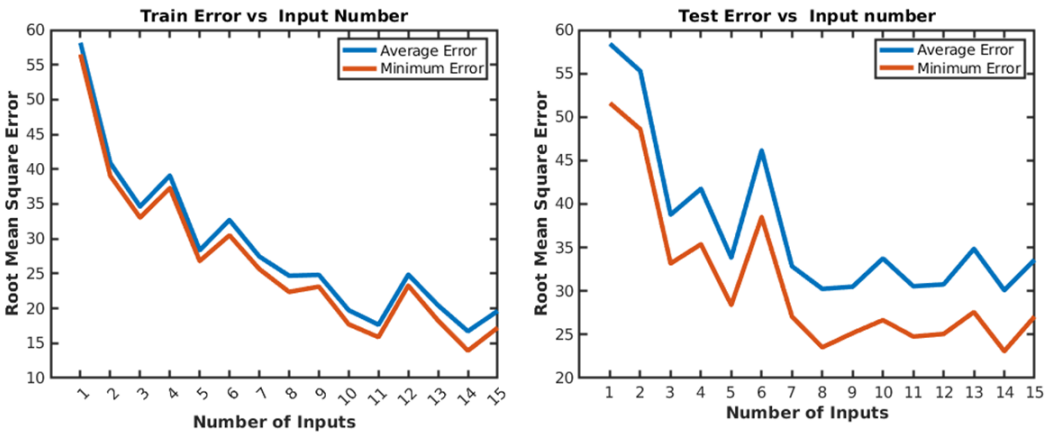

Figure 3.2. Minimum and average train and test errors vs number of inputs for random subset of features.

It can be seen that 8 and 14 features give lower errors for test sets in Figure 3.2. Hence, 10 feature set are represented that gives minimum test errors for regression tree method in Table 3.8. Then, these feature sets are computed with artificial neural networks and results are compared in Table 3.9. Number of neurons are tuned for each case. Case 6 with 14 features give lower error with 16 neurons. The test error and feature set is qualified, yet, lower test errors can be achieved. Hence, random forest feature importance also implemented to exploration of important features.

Table 3.8. Feature Sets.

Case No	**Train RMSE**	**Test RMSE**	**Feature Sets**
1	16.402	23.091	ElectMG HeatT SSA IDIG Nconcentration Oconcentration Pore Volume Electrolyte Type Separator PrepMG Salt Anion Volume Salt Cation Volume Solvent Dipole Moment Binder Concentration
2	18.158	23.278	ElectMG HeatT SSA IDIG Oconcentration Pore Volume Electrolyte Type Separator Current Collector Normalized Scan Rate Absolute Potential Window Salt Cation Volume Solvent Dipole Moment Binder Concentration

Table 3.8. Feature Sets. (cont.)

Case No	**Train RMSE**	**Test RMSE**	**Feature Sets**
3	16.151	23.471	ElectMG HeatT IDIG Nconcentration Oconcentration Pore Volume Electrolyte Concentration Separator PrepMG Current Collector Normalized Scan Rate Absolute Potential Window Salt Anion Volume Salt Cation Volume
4	15.937	23.527	ElectMG HeatT SSA Nconcentration Oconcentration Pore Volume Electrolyte Type Electrolyte Concentration PrepMG Current Collector Normalized Scan Rate Absolute Potential Window Solvent Volume Salt Anion Volume

Table 3.8. Feature Sets. (cont.)

Case No	**Train RMSE**	**Test RMSE**	**Feature Sets**
5	16.346	23.847	ElectMG HeatT Nconcentration Pore Volume Electrolyte Type PrepMG Current Collector Normalized Scan Rate Absolute Potential Window Solvent Volume Salt Anion Volume Salt Cation Volume Solvent Dipole Moment Binder Concentration
6	16.671	23.875	ElectMG HeatT SSA IDIG Nconcentration Pore Volume Electrolyte Concentration Current Collector Separator Normalized Scan Rate Absolute Potential Window Salt Cation Volume Solvent Volume Binder Concentration

Table 3.8. Feature Sets. (cont.)

Case No	**Train RMSE**	**Test RMSE**	**Feature Sets**
7	16.738	24.013	HeatT SSA IDIG Nconcentration Oconcentration Pore Volume Electrolyte Type Current Collector Normalized Scan Rate Absolute Potential Window Solvent Volume Salt Cation Volume Solvent Dipole Moment Binder Concentration
8	15.510	24.085	ElectMG HeatT SSA IDIG Nconcentration Oconcentration Electrolyte Type Electrolyte Concentration PrepMG Separator Normalized Scan Rate Salt Anion Volume Salt Cation Volume Solvent Volume

Table 3.8. Feature Sets. (cont.)

Case No	**Train RMSE**	**Test RMSE**	**Feature Sets**
9	17.121	24.109	ElectMG HeatT IDIG Nconcentration Oconcentration Electrolyte Type Electrolyte Concentration PrepMG Separator Current Collector Absolute Potential Window Salt Anion Volume Solvent Dipole Moment Binder Concentration
10	15.650	24.303	ElectMG HeatT SSA Oconcentration Pore Volume Electrolyte Type Electrolyte Concentration PrepMG Separator Normalized Scan Rate Salt Anion Volume Salt Cation Volume Solvent Dipole Moment Binder Concentration

Table 3.9. Results of artificial neural networks for randomly created subset features.

CASE	R^2			MSE	RMSE	Neuron
	Train	Validation	Test	(Test)	(Test)	Number
1	0.956	0.941	0.936	546.22	23.371	20
2	0.960	0.929	0.940	520.35	22.811	19
3	0.949	0.918	0.914	612.97	24.758	15
4	0.952	0.947	0.929	567.53	23.823	16
5	0.934	0.924	0.929	507.18	22.520	15
6	0.970	0.957	0.952	374.71	19.357	16
7	0.922	0.901	0.915	711.13	26.667	11
8	0.958	0.904	0.915	595.42	24.401	13
9	0.914	0.913	0.882	901.66	30.027	16
10	0.948	0.901	0.904	754.82	27.474	10

3.2.1.3. Random Forest Feature Importance. Feature selection and relevance can assist perform and interpreting machine learning algorithms efficiently. In machine learning, ensemble algorithms have had accomplishments by combining numerous weak learners to create powerful learner. Random forest is one of the successful ensembles that use decision trees as base algorithm and, information gain to split nodes. The random forest model generates randomly selected features that allow careful examination of possible subsets of features. Furthermore, it uses bagging model that uses a certain part of training data to construct model. The out-of-bag data is used as a testing set to check the model. Overall, random forest method is self-checking for feature selection and their relevance. Feature selection methods should examine subset of features instead of individually to achieve satisfactory dimension reduction [62]. For these reasons, random forest predictor importance was implemented to assess the quality of features for a robust model. Random forest algorithm is computed to find importance of features to predict capacitance. Importance of predictors are shown in Figure 3.3.

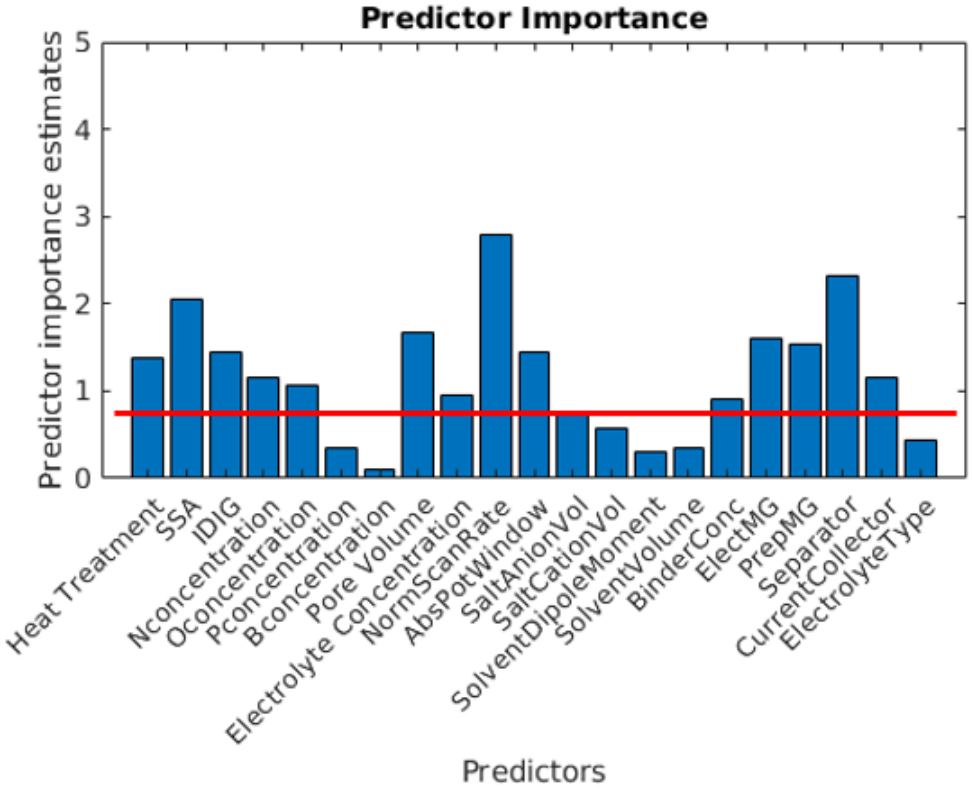

Figure 3.3. Random Forest predictor importance.

Bconcentration, Pconcentration, Solvent volume, Solvent dipole moment, Electrolyte type, Salt anion volume and Salt cation volume do not have significant impact on dataset according to random forest feature importance method. Also, obtained parameter set shows rational parameter set to design optimum supercapacitor model since these features encompass electrode/electrolyte material information and operation conditions. The results of trial/error method supports the random forest feature importance model for several predictors such as ElectMG, SSA, PrepMG, Absolute potential window and Separator. It can be noted that 12 features in Case 6 verify the results of Random Forest feature importance model and trial/error. Random Forest feature selection can be considered as efficient feature selection method for supercapacitor dataset. Therefore, 14 features are selected to build machine learning models by using random forest feature importance model. These features are HeatT, SSA, IDIG, Nconcentration, Oconcentration, Pore Volume, Electrolyte Concentration, Normalized Scan Rate, Absolute Potential Window, Binder Concentration, ElectMG, PrepMG,

Separator and Current Collector. Results of machine learning algorithms for 14 features are represented in Table 3.10.

Table 3.10. Results of machine learning algorithms with selected features from Random Forest predictor importance method.

Train Ratio	**80%**	
Algorithms	**Train**	**Test**
Linear Regression	39.499	40.852
Lasso Regression	40.208	40.448
Regression Tree	19.827	30.715
Random Forest	18.563	26.286
Gradient Boosting	23.299	26.355
ANN	23.388	27.313

3.3. Statistical Learning Models

3.3.1. Parameter Tuning

Hyperparameters are various variables in machine learning algorithms that can improve the behavior of learning algorithms. The hyperparameters should be tuned before the main process for each learning model. Hyperparameter tuning can be proceeded manually and automatically. Grid search is effective way for machine learning models that have two tunable parameters. Random search or Bayesian hyperparameter optimization can be more suitable if there are more variable. In this work, grid search with cross-validation is used to adjust hyperparameters by manually [14,38]. Parameters are selected according to cross-validation results. Optimum parameter that gives minimum average error is selected. Tuning parameters are shrinkage parameter for penalized regression (lasso), the minimal number of observations per tree leaf and complexity parameter for regression tree, number of variables available for splitting at

each tree node for random forest. Grid search is performed for gradient boosting and artificial neural networks. Learning rate and number of learning cycles for gradient boosting, number of neurons and number of hidden layers for artificial neural networks are tuned by manually.

3.3.2. Linear Regression

Linear regression method is applied as a simple machine learning algorithm to predict capacitance. R^2 is found that 0.57 and 0.55 for training and testing set that means there is quite low fitting. The actual and predicted values can be seen from Figure 3.4. RMSE is 39.48 F/g and 41.57 F/g for train and test dataset, respectively. Therefore, linear regression is not promising method to predict capacitance due to the high errors and low fitting. In order to obtain more accurate results, different methods should be considered.

3.3.3. Lasso Regression

Penalized regression approach with lasso penalty is applied for different lambdas as tuning parameters. λ is 1.37 that gives a more regularized model (lambda.1SE) that can be seen in Figure 3.5. RMSE values are 40.86 F/g and 42.19 F/g for training and testing set, respectively. R^2 is obtained as 0.54 and 0.53 for training and testing set. Predicted and actual capacitances can be seen from Figure 3.6. Also, linear and lasso regression have similar patterns and low fitting. The cross-validation errors are not satisfactory for linear and lasso regression algorithms, so they are not useful models to predict capacitance.

3.3.4. Regression Tree

Regression tree is performed to predict capacitance. The minimal number of observations per tree leaf is considered for model evaluation with 5, 10 and 15. The best result is found by using 5 by manual tuning.

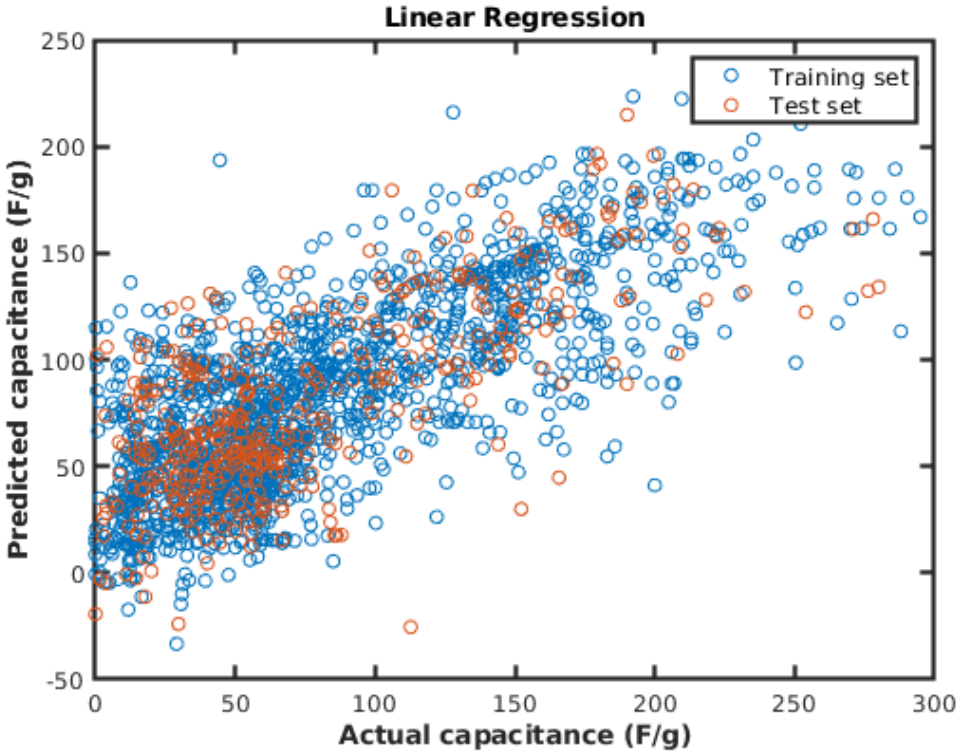

Figure 3.4. Predicted vs actual capacitance values for Linear Regression (80% training set).

RMSE of training and testing dataset are determined as 21.42 F/g and 30.32 F/g. It's much better than linear and lasso regression. However, further attempts are needed to achieve more satisfactory results. R^2 is achieved as 0.87 and 0.76 for train and dataset. Real and predicted capacitances by regression tree are represented in Figure 3.7.

3.3.5. Random Forest

Random forest model is implemented to get more efficient predictive performance than regression tree. The hyperparameter that is used for model evaluation is the number of variables randomly sampled. The number of variables randomly sampled is tuned for 3, 5, 7 and 10. When the number of features is increased, the error is slightly decreased. Therefore, the final model is built for 10.

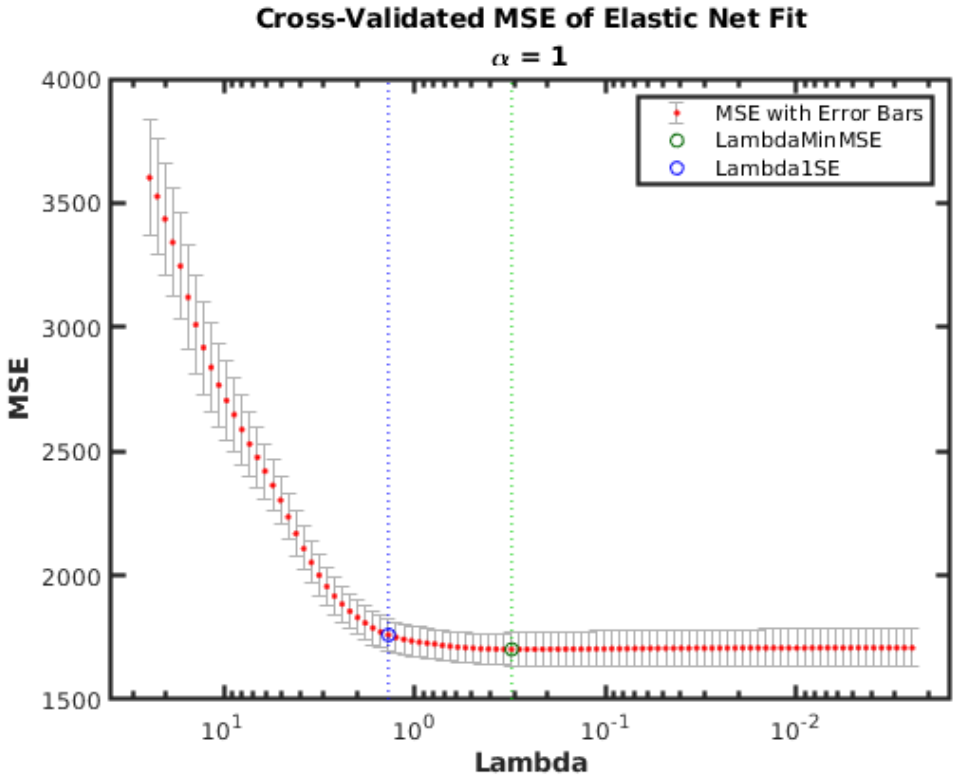

Figure 3.5. Cross-validated mean square error of elastic net fit with $\alpha = 1$.

RMSE values of training and test set are 20.30 F/g and 27.97 F/g, respectively. It is expected that random forest resulted in low error than regression tree since the random forest is an improved form of regression tree algorithm. It can be observed from Figure 3.8. Also, random forest model possess impressive fitting with 0.88 and 0.79 as R^2 of training and testing data.

3.3.6. Boosting Method

Gradient boosting is computed as another ensemble method. In this approach, parameters to be tuned are learning rate (shrinkage) and number of learning cycles. Grid search is used as manual hyperparameter tuning. The optimal learning rate and number of learning cycles are found as 100 and 0.1 by using cross-validated results. RMSE is 23.32 F/g and 29.23 F/g for the train and test set. Correlation between actual and predicted capacitance is demonstrated in Figure 3.9.

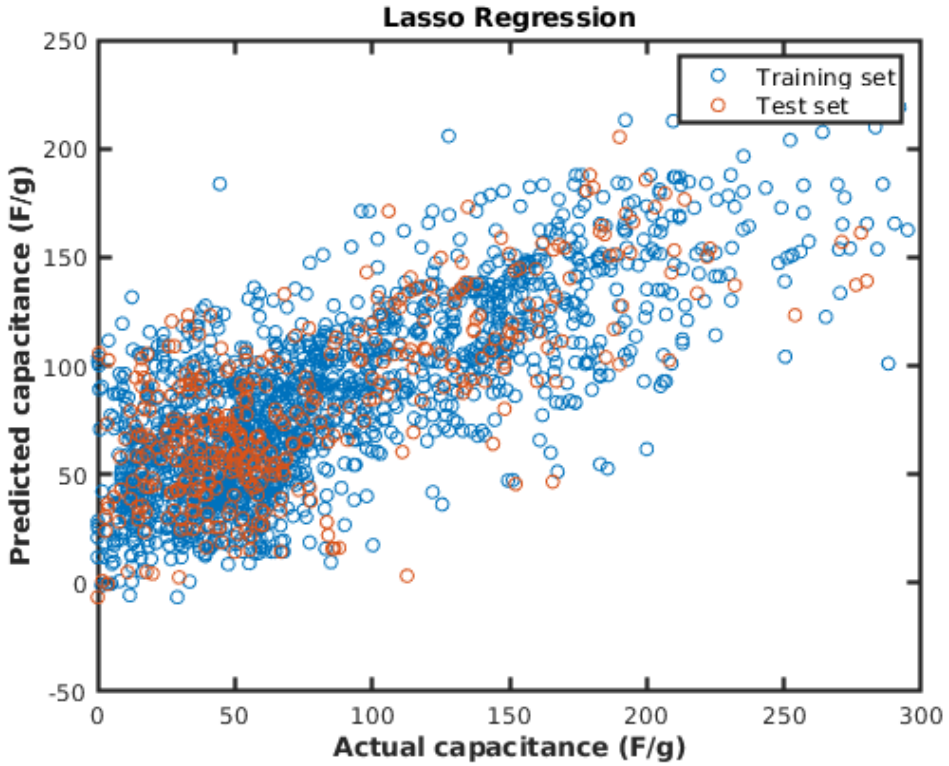

Figure 3.6. Predicted vs actual capacitance values for Lasso Regression (80% training set).

R^2 values are 0.85 and 0.78 for training and testing data, respectively. Random forest shows better performance than gradient boosting as an ensemble algorithm.

3.3.7. Artificial Neural Networks

In this study, Neural Network algorithms are implemented with MATLAB's Deep Learning Toolbox. 10-fold cross-validation method is used to find the best structure that provides a higher fitting and lower errors. The selection of activation function depends on the structure of data and target variables. Sigmoidal functions such as hyperbolic tangent and logistic sigmoid are widely used as activation functions in neural networks. Also, they are highly related since $tanh(x) = 2\sigma(2x) - 1$ [38, 39].

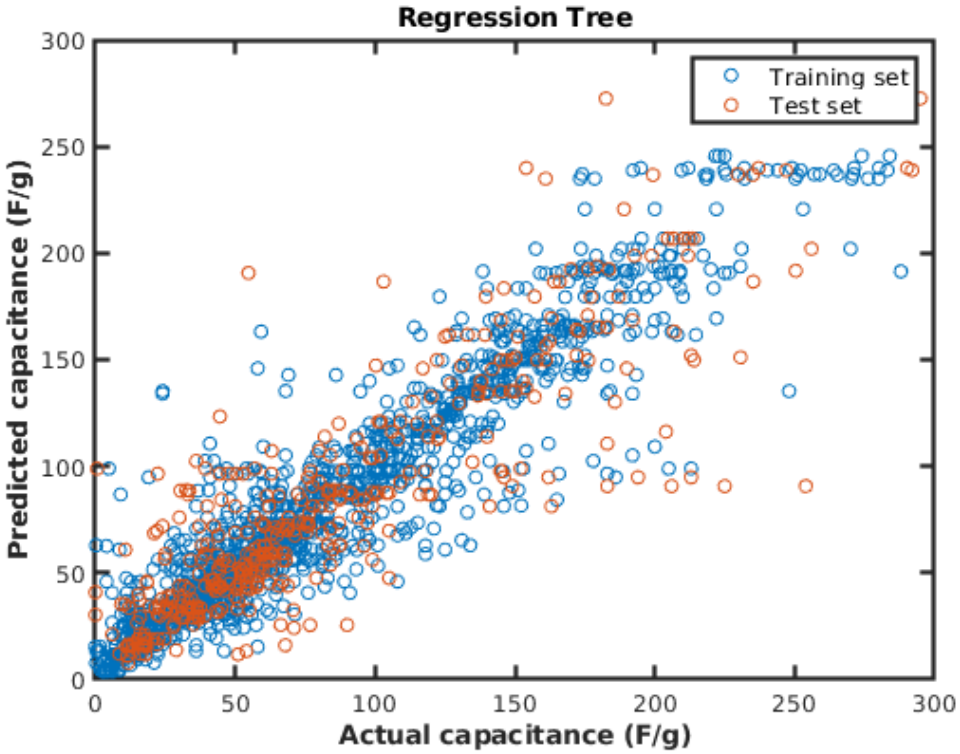

Figure 3.7. Predicted vs actual capacitance values for Regression Tree (80% training set).

Grid search is used to tune activation function and number of neurons. The sigmoid and hyperbolic tangent activation function is used as a transfer function to examine the effect of the activation function. Also, number of neurons is tuned for artificial neural network that can be seen in Table 3.11 and Table 3.12. These results best folds of 10-fold cross-validation that gives minimum train and test errors. It is observed that 16 and 20 neurons for artificial neural networks show better performance with sigmoid and hyperbolic tangent activation function, respectively. Hence, sigmoid activation function is utilized as activation function of ANN model. RMSE values are 14.20 F/g and 15.79 F/g for the train and test set, respectively. Moreover, R^2 values are achieved as 0.94 for the training set and 0.93 for the testing set. Predicted and real capacitances can be seen in Figure 3.10.

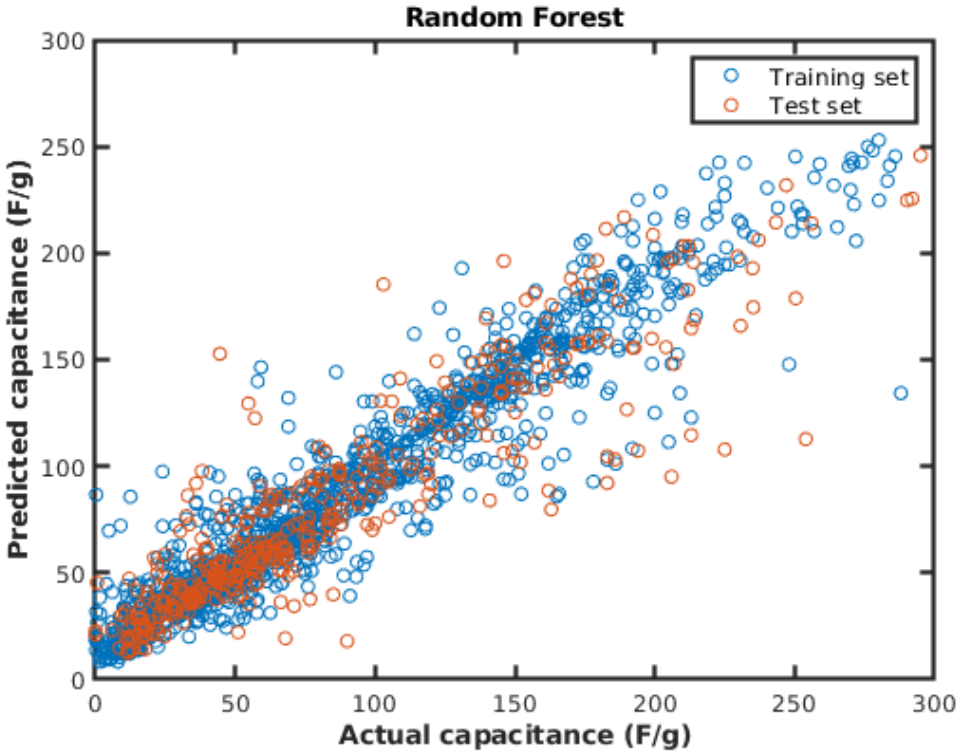

Figure 3.8. Predicted vs actual capacitance values for Random Forest (80% training set).

It is worth to mention that highest R^2 is achieved 0.96 by using 2 hidden layer and 50 neurons and the lowest RMSE is 18.13 F/g in the literature to the best of our knowledge. In the light of such information, ANN shows excellent performance in terms of low error and high fitting with the low number of neurons (16) and one hidden layer in this work.

3.3.8. Comparison of Different Machine Learning Models

Linear regression, lasso regression, regression tree, random forest, gradient boosting, and artificial neural networks are used to construct models that can predict capacitance. The results of different machine learning algorithms are examined for model selection. Machine learning algorithms are computed for different training ratios to understand effect of number of data points in the training set.

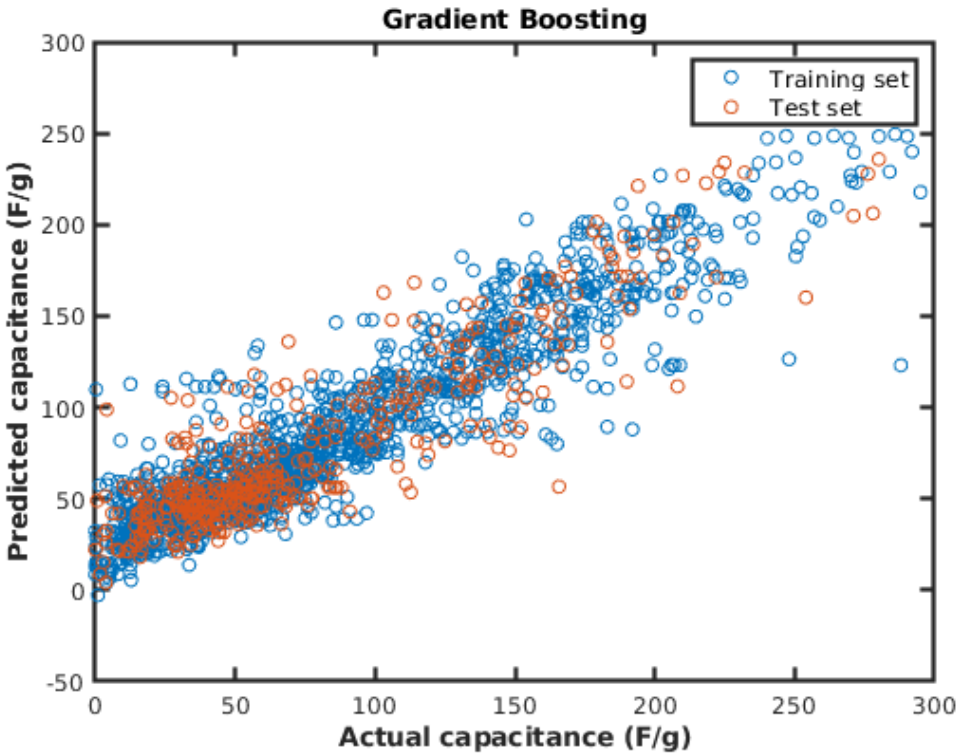

Figure 3.9. Predicted vs actual capacitance values for Gradient Boosting (80% training set).

The results are represented in Table 3.13. When the number of data points is decreased to build the model, test errors are increased. Also, root mean square errors of machine learning algorithms for cross-validated models are presented in Table 3.14. Artificial neural networks shows better performance than other algorithms. Therefore, 16 neurons and sigmoid activation function are selected to build neural network model and used as the fitness function of genetic algorithm.

3.3.9. Artificial Neural Networks-Genetic Algorithm Hybrid Model

3.3.9.1. Optimal Solution of Hybrid ANN-GA Model. The hybrid ANN-GA model is introduced to optimize the capacitance of supercapacitors according to several constraints. The output of artificial neural networks is used as a fitness function of genetic algorithm.

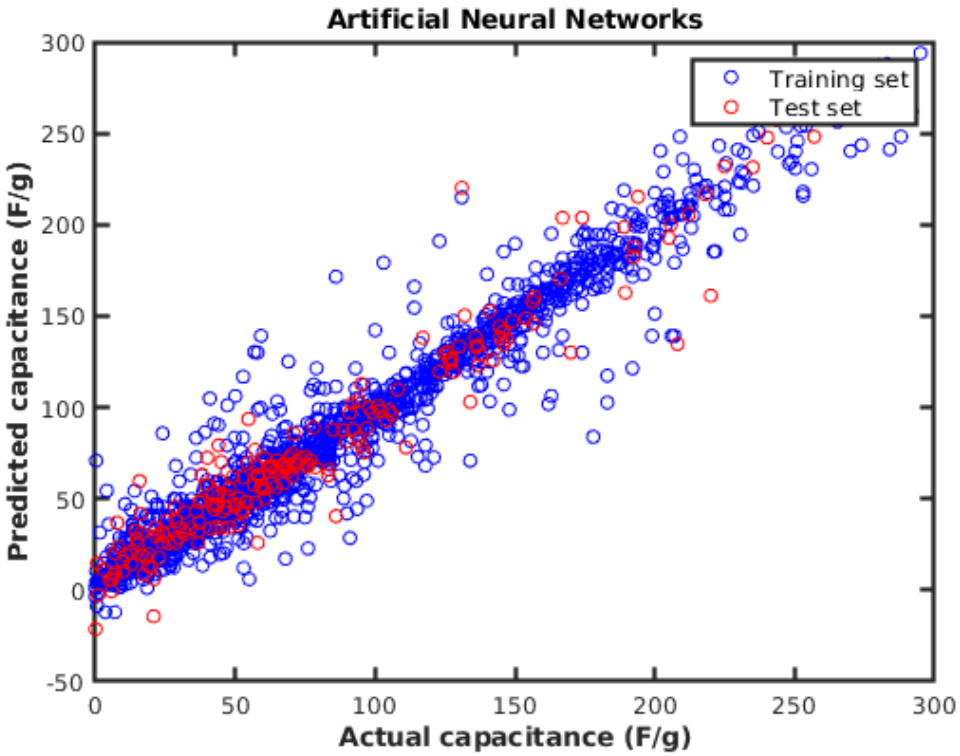

Figure 3.10. Predicted vs actual capacitance values for Artificial Neural Networks (80% training set).

Genetic algorithm is a practical tool to explore search space and increase input space. However, constraints are needed to lead more satisfactory results. Linear equality constraints are specified for categorical features, linear inequality constraints are used to express relationships of two predictors by using big-M method, and side constraints are utilized to define lower and upper limits. The constraints are specified by using literature information and relationship of features from the dataset. Also, process parameters are defined for genetic algorithm optimizer. Inıtial population is 500 and maximum number of generations is 100, function and constraint tolerance is 10^{-6}. Crossover fraction is specified as 0.8, elite members that are transferred next generation is 25. The "ga" function of MATLAB software is used as a solver for genetic algorithm and it solves minimization problems by default. Therefore, the objective function is multiplied by -1 to obtain the maximization problem.

Table 3.11. Effect of neuron number for artificial neural networks with sigmoid activation function (80% training set).

Number of Neurons	R^2	
	Train	**Test**
10	16.271	20.391
11	16.633	19.501
12	17.064	19.614
13	18.924	19.610
14	15.069	19.369
15	15.084	16.505
16	14.205	15.795
17	16.513	19.026
18	14.702	18.769
19	16.219	19.261
20	14.912	16.903

Artificial neural networks-Genetic algorithm based hybrid model is used to obtain optimal capacitance and parameter set. Literature review is carried out to examine the parameter set that is obtained from output of ANN-GA model. In this context, graphene/carbon nanotubes (CNTs) based materials are investigated, and the feasibility of the parameter set is discussed. The maximum capacitance is achieved as 1012 F/g by ANN-GA hybrid model with 10^{-6} function and constraints tolerance that is represented in Figure 3.11. The parameter set includes several features that are described for supercapacitors such as heat treatment temperature, specific surface area, pore volume, ID/IG ratio, electrolyte concentration, potential window, binder concentration, separator, nitrogen concentration, oxygen concentration, and current collector. The optimal parameter set that is obtained from ANN-GA can be seen in Table 3.15.

Table 3.12. Effect of neuron number for artificial neural networks with hyberbolic tangent activation function (80% training set).

Number of Neurons	R^2	
	Train	**Test**
10	16.610	19.992
11	16.462	19.905
12	17.373	19.891
13	15.765	19.486
14	15.478	19.743
15	16.123	19.661
16	15.500	18.872
17	17.154	18.749
18	15.442	19.715
19	14.759	18.835
20	14.966	16.417

Table 3.13. Results of ML approaches with different training ratios for all data without cross-validation.

Train Ratio	**80%**		**70%**		**50%**		**20%**	
Algorithms	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
Linear Regression	38.696	39.333	39.086	38.413	38.796	39.446	37.835	45.605
Lasso Regression	39.479	39.474	39.634	38.728	40.189	40.566	40.086	43.478
Regression Tree	19.927	31.806	20.594	30.809	22.382	32.378	26.926	43.601
Random Forest	17.952	25.753	18.978	26.475	20.892	28.135	16.580	35.683
Gradient Boosting	22.635	25.949	22.546	25.793	21.440	28.135	16.580	35.683
ANN	18.055	25.020	22.068	27.695	24.879	34.785	27.965	39.899

Table 3.14. Results of ML approaches for cross-validated models.

Algorithms	**Train**	**Test**
Linear Regression	39.478	41.575
Lasso Regression	40.858	42.193
Regression Tree	21.425	30.318
Random Forest	20.301	27.971
Gradient Boosting	23.323	29.231
ANN	14.205	15.795

Table 3.15. Optimal solution of ANN-GA hybrid model.

Numerical Features					
	Heat Treatment	**SSA**	**ID/IG**	**Nitrogen Conc.**	**Oxygen Conc.**
minimum	25	0.99	0.35	0	0
optimal	868	2498	1.5	6.5	10.0
maximum	1000	3523	4	20.6	28.8
	Pore Volume	**Electrolyte Conc.**	**NSR**	**APW**	**Binder Conc.**
minimum	0.003	0.5	0	0.4	0
optimal	2.5	4.8	0.5	4.7	8.4
maximum	4	8.1	1	4.8	15
Categorical Features					
ElectMG	Graphene/CNT				
PrepMG	Chemical Vapor Deposition				
Separator	Cellulose				
Current Collector	Au/Pt				

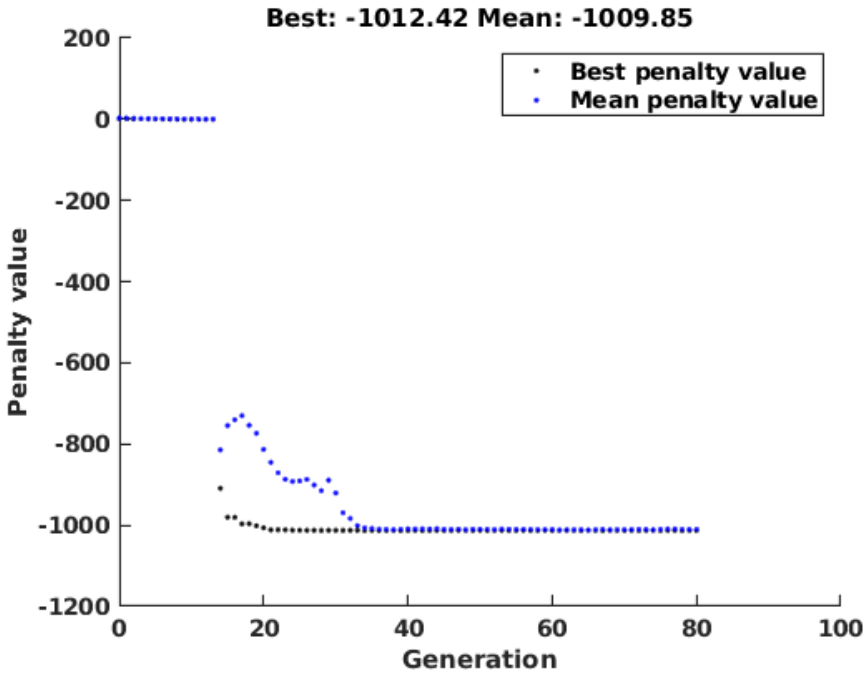

Figure 3.11. Optimal capacitance of ANN-GA hybrid model.

The electrode material is found as N-doped graphene/carbon nanotubes (CNTs) hybrid material for optimal solution. Hybrid carbon materials and ionic liquids as electrolyte material provide unique properties for high performance supercapacitors. Activated carbon is widely used as electrode material in supercapacitors. However, the presence of micropores in high surface area of activated carbon limits the accessibility of electrode and electrolyte which results in low specific capacitance. Nanomaterials such as carbon nanotubes (CNTs) have drawn attention since they possess large surface area, chemical stability, tunable porosity, and increased electrode/electrolyte interaction. Also, graphene shows great specific capacitance performance due to the high surface area. However, the tendency for agglomeration of graphene layers reduces the performance of graphene as an electrode. The graphene/CNTs hybrid structure is designed to achieve high-performance supercapacitors and complementary electrode material [63].

Carbon nanotubes such as single-walled CNTs (SWCNTs), multi-walled CNTs (MWCNTs) and double-walled CNTs (DWCNTs) can be prepared by using chemical vapor deposition (CVD) methods at high temperatures. The better catalyst is Fe/MgO to synthesize SWCNTs and DWCNTs at 900-1000°C by decomposition of methane [64]. Also, graphene can be produced via chemical vapor deposition at 800-1000°C. Several CVD methods are attempted to produce graphene/CNTs material such as one-step CVD method and catalyst-floating CVD method [64]. Therefore, the optimal preparation method of electrode (PrepMG) and heat treatment of electrode material (HeatT) can be employed to graphene/CNTs supercapacitors.

Electrode materials that exhibit high specific surface area are preferable in supercapacitor applications. Electrolyte ions can be diffused into the available pores easily thanks to the large surface area [9]. The specific surface area of optimal solution is 2498 m^2/g for electrode material. For graphene/CNTs based materials, the specific surface area is achieved as 2000-2600 m^2/g [65]. Hence, the optimal solution applicable for graphene/CNTs electrode. The pore structure of electrode is as important as the specific surface area for improved capacitance. A porous structure is required to enable the accessibility of electrode ions during the charge/discharge process. If the electrolyte ions cannot be diffused completely into electrode material, performance of the supercapacitor is reduced despite high surface area [9]. Pore size is 2.5 cm^3/g for optimal design of supercapacitor. The effect of SSA and pore volume on capacitance can be observed in Figure 3.12. Design maps are used to examine relationship of two features and their effect on capacitance. For this reason, certain two features are divided into equal grids within the range of their minimum and maximum values, and the remaining features are fixed to the optimal values for ANN model.

The pore structure and size distribution are crucial factors to increase the capacitive performance of supercapacitors. Specific surface area of electrode material determines the magnitude of capacitance, whereas pore structure plays an essential role in electron transfer and ion transport. Therefore, optimization of specific surface area and pore volume lead to enhanced performance in supercapacitors [66].

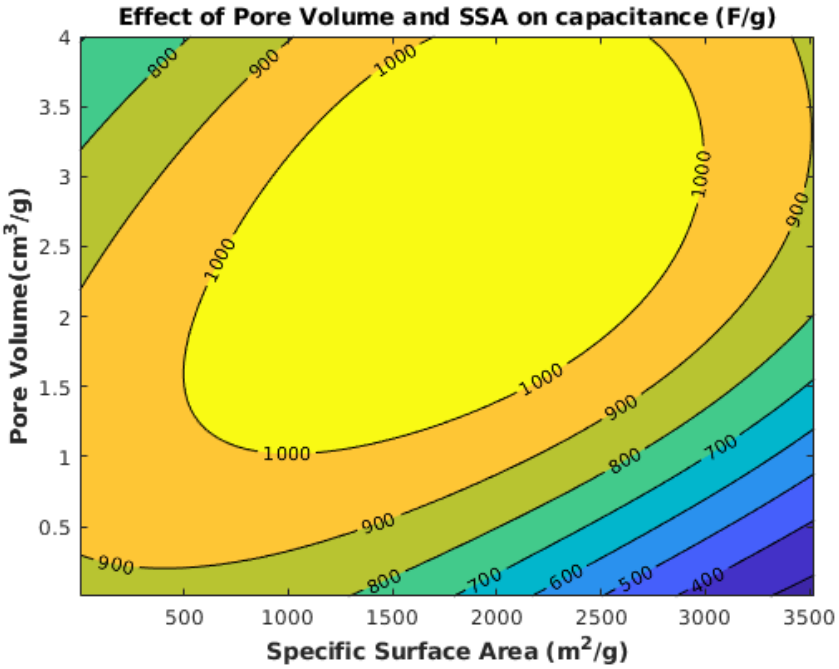

Figure 3.12. Effect of SSA and Pore Volume on capacitance.

It can be seen that from Figure 3.12, optimum specific surface area and pore volume enables high capacitance.

Temperature is the key parameter that defines physicochemical properties of carbon materials. The specific surface area and yield of carbonized compounds increase with the enhancement of activation temperature due to the promoting pore formation. However, higher temperatures can degrade the porous structure of carbon materials which results in decreased surface area [9, 67, 68]. The influence of heat treatment for specific surface area and pore volume is shown in Figure 3.13 and Figure 3.14, respectively. It can be seen that optimum activation temperature is needed to achieve desirable capacitance.

Moreover, temperature affects the degree of carbonization. Heat treatment enhances the accessible surface area of the electrode to ease the interaction between electrolyte ions and electrode material.

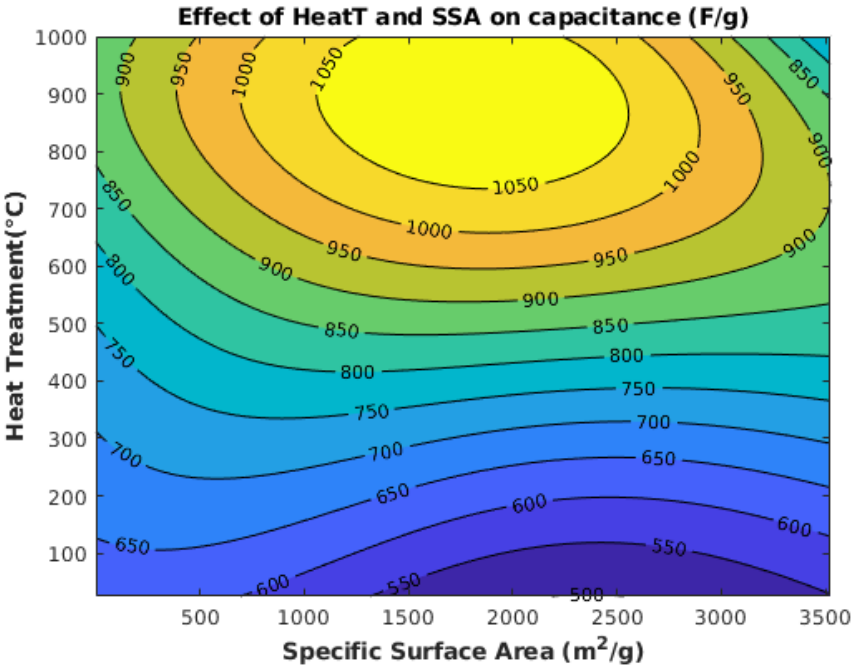

Figure 3.13. Effect of HeatT and SSA on capacitance.

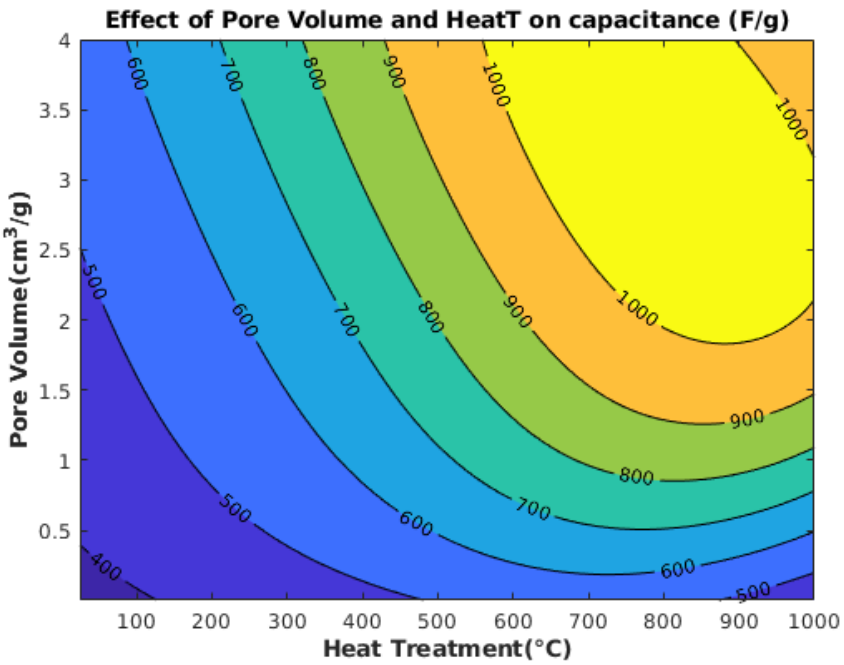

Figure 3.14. Effect of HeatT and Pore Volume on capacitance.

Also, heat treatment eliminates oxygenated functional groups to a certain extent according to activation temperature since high oxygen functionalities may decrease conductivity. However, heat treatment may lead to some undesirable textural changes. At high temperatures, the porosity of electrode material can decrease which resulted in decreased electrochemical performance.

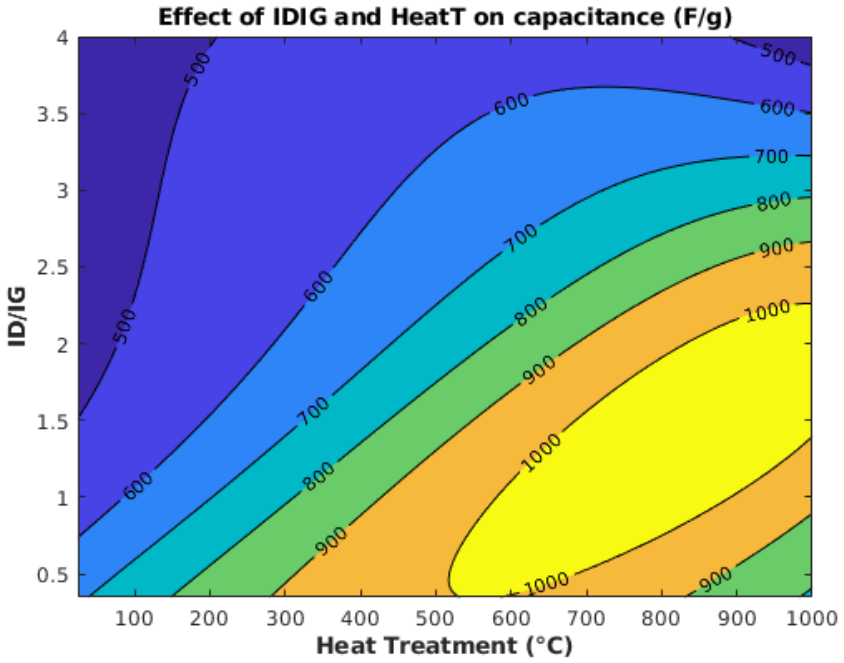

Figure 3.15. Effect of HeatT and ID/IG on capacitance.

Also, additional materials such as binder may decompose at high temperatures and degrade the electrode structure [69]. Raman spectrum can be used to identify the graphitic structure of carbon materials. The higher ID/IG ratio leads high graphitic nature that provides good conductivity. For carbon materials, varying carbonization temperatures create different ID/IG ratios. This behavior can be observed in different carbon structures such as graphene, carbon nanospheres, carbon nanotubes, etc. The results indicate that temperature is a crucial factor in the electrochemical performance of supercapacitors [9,70,71]. A similar pattern of HeatT and ID/IG for graphene/CNT material can be seen in Figure 3.15. Although the graphitization increases electrical conductivity, porosity decrease at higher temperatures. This leads to high defects and

disorders which result in poor conductivity. It can be observed that higher ID/IG ratio corresponds to lower porosity from the Figure 3.16.

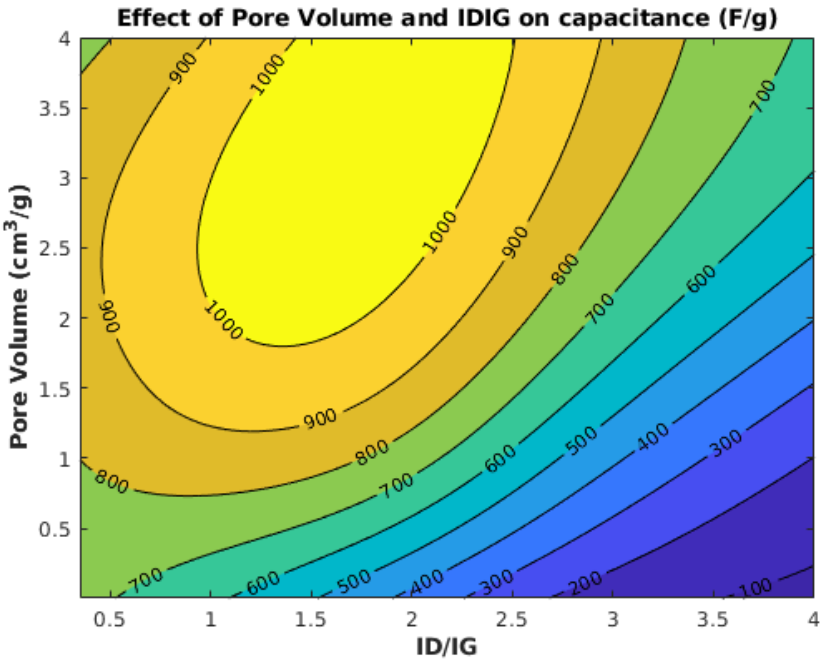

Figure 3.16. Effect of ID/IG and Pore Volume on capacitance.

To further enhance performance, introducing heteroatoms into electrode material is a powerful method. Among heteroatoms, nitrogen has been widely used due to the wettability and pseudocapacitance effect. Nitrogen doping is applied with graphene/CNTs materials and performance enhancement is acquired [72, 73]. It is observed that ultra-high capacitances are achieved by nitrogen-doped carbon materials. Ye *et al.* reported nitrogen-doped carbon nanospheres (CNs) that possess a high specific capacitance of 627 F/g in aqueous electrolyte [74]. Li *et al.* obtained a high-performance supercapacitor with the hybrid electrode material. Graphene/CNs hybrid composite has excellent specific capacitance as 720 F/g with 12.19% nitrogen content [75]. Also, graphene-based supercapacitors exhibit high capacitances owing to nitrogen doping. Sun *et al.* proposed nitrogen-doped three-dimensional (3D) graphene as an electrode material. The specific capacitance of 7.81% nitrogen-doped 3D graphene is achieved as 788 F/g [76]. Therefore, nitrogen-doped carbon materials are promising

candidates that effectively enable ultra-high capacitance. In general, oxygen is inserted to mediate nitrogen since both have higher pseudocapacitance effect together [77]. The optimal nitrogen and oxygen concentration of ANN-GA model are 6.5% and 10.0%, respectively. Also, the ratio of nitrogen and oxygen amount is feasible according to literature information.

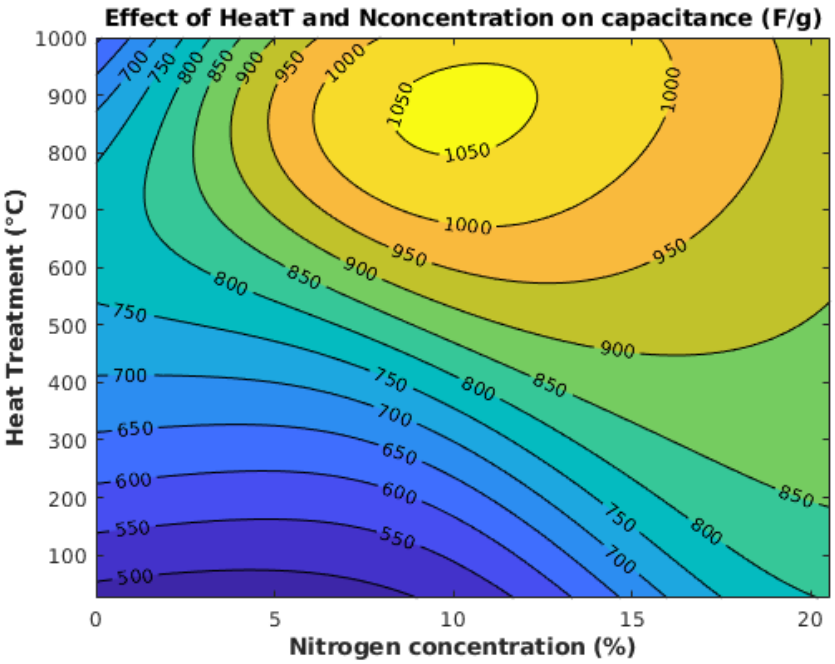

Figure 3.17. Effect of Nitrogen concentration and HeatT on capacitance.

Nitrogen and oxygen concentration changes according to different activation temperatures. The effect of activation temperature and nitrogen/oxygen amount can be observed in Figure 3.17 and Figure 3.18, respectively. In addition, Ye *et al.* found that the higher nitrogen content is obtained at 900°C with respect to 800°C, whereas the oxygen concentration at 900°C is lower than 800°C for nitrogen-doped carbon nanospheres. Also, the optimal amount of oxygen and nitrogen should be doped to achieve high capacitance since the excessive doping result in poor conductivity. Ilnicka *et al.* reported that nitrogen doping should be kept in the optimum range to achieve satisfactory specific capacitance [74, 78]. The doping of nitrogen improves the wettability and conductivity of carbon materials. The introduction of electron-rich nitrogen into the carbon

networks increases the number of free electrons of carbon materials that resulting in enhanced electrical conductivity. In general, oxygen functionalities come along with nitrogen doping. The presence of oxygen-containing groups contributes to redox reaction that leads to enhanced capacitance. The combined pseudocapacitive effect of nitrogen and oxygen improves the performance of the supercapacitor. Also, a certain level of nitrogen doping offers transport paths for highly efficient ion transfer. However, excessive nitrogen may damage the structural and chemical properties of carbon material [79].

The effect of nitrogen doping on enhanced capacitance is obvious, yet, there is a lack of investigation on the underlying mechanism. Jeong *et al.* reported that nitrogen-doped graphene has 4 times higher capacitance compared to undoped graphene. It is claimed that the increased capacitance may arise from improved binding interactions between electrolyte ions and nitrogen dopants in the basal plane [80]. Also, Paek *et al.* observed that doping type affects the capacitance. The substitutional nitrogen and trimerized pyridine type nitrogen configuration are considered to observe the change in capacitance. The substitutional nitrogen doping demonstrates greater enhancement with regard to capacitance. It is clear that both N-doped models possess higher capacitances than undoped graphene [81]. Jonna *et al.* reported a substantial increase in capacitance with nitrogen doping for graphene electrodes. The specific capacitance of graphene oxide and nitrogen-doped graphene are 14 F/g and 118 F/g, respectively [82]. Another study is proposed by Sun *et al.* for nitrogen-doped graphene material. It is reported N-doped graphene enables much higher capacitance than reduced graphene oxide. The N-doped graphene achieves 326 F/g with high nitrogen content (10%), whereas reduced graphene oxide has 203 F/g as specific capacitance at the same operating conditions [83].

Furthermore, it is observed that nitrogen and oxygen functionalities improve the graphitic structure of carbon materials. The higher nitrogen/oxygen doping corresponds to higher ID/IG ratio. The effect of nitrogen and oxygen doping on ID/IG for electrochemical performance can be seen in Figure 3.19 and Figure 3.20.

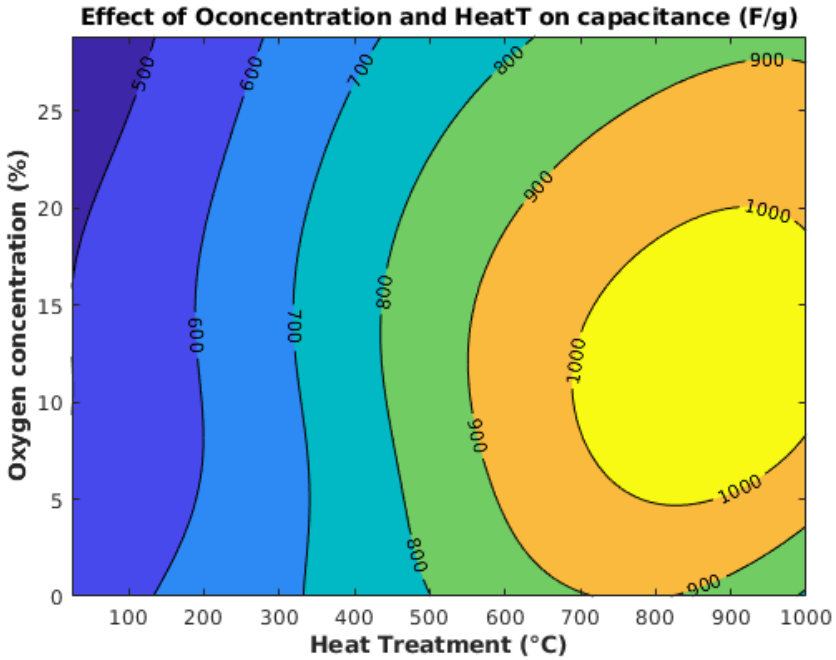

Figure 3.18. Effect of Oxygen concentration and HeatT on capacitance.

It can be observed that appropriate content of nitrogen and oxygen leads to optimum ID/IG ratio that results in enhanced specific capacitance. Lin *et al.* observed the effect of nitrogen doping on graphitic nature for graphene/CNTs supercapacitors. The ID/IG ratio of nitrogen-doped graphene/CNTs is 1.76, whereas the ID/IG is 1.57 for graphene/CNTs. The specific capacitance is 10 F/g and 168 F/g for graphene/CNTs and nitrogen-doped graphene CNTs, respectively. Therefore, nitrogen doping contributes the graphitic structure and increase specific capacitance [84]. Also, the effect of the carbon and oxygen (C/O) ratio is investigated for electrochemical performance of supercapacitors. Cao *et al.* reported different oxygen content for graphene sheets to explore the effect of oxygen amount on supercapacitors. Raman spectroscopy is performed to get an idea of the impact of oxygen functional groups and defects. The ID/IG ratios were determined as 0.96, 0.94, 0.93 for the corresponding C/O ratio of 4.1, 8.3, and 13.5, respectively. Also, the highest specific capacitance was obtained for the sample that has higher oxygen functional groups (C/O ratio is 4.1). The results indicate that high oxygen content contributes to specific capacitance by providing ad-

ditional pseudocapacitance. [85,86]. Pope *et al.* also presented varied ID/IG ratios and specific capacitance according to the content of oxygen. However, very high oxygen content reduces the specific capacitance by the reason of switching from an electrically conductive state to insulating [85, 87].

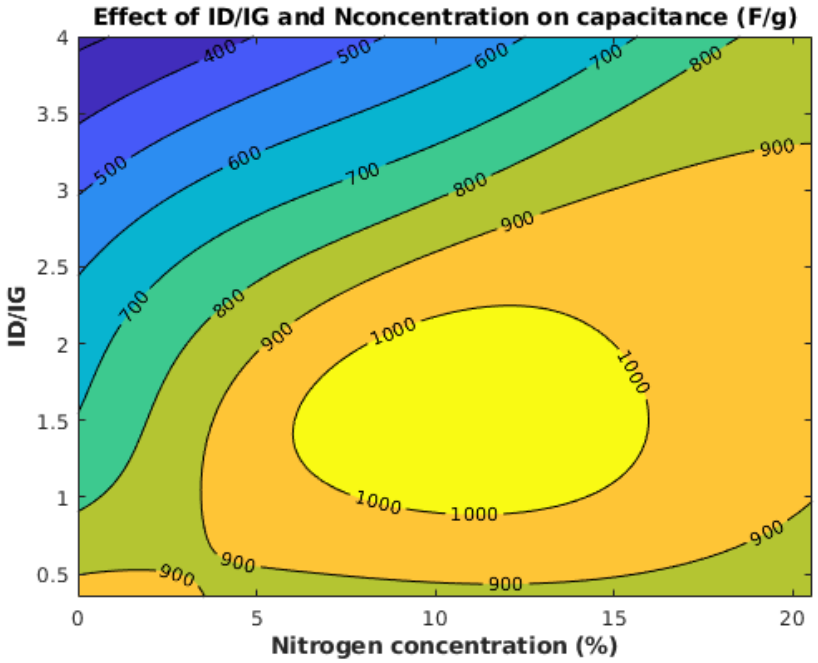

Figure 3.19. Effect of Nitrogen concentration and ID/IG on capacitance.

Supercapacitors store the energy at the electrode/electrolyte interface via physical charge storage. The relationship between the specific surface area of electrode materials and accessible electrolyte ions determines the specific capacitance. Electrolyte material selection is a key parameter to increase the capacitive performance of supercapacitors. Until now, there has not been encountered a perfect electrolyte that satisfies all the needs of the electrochemical device. The optimization and compatibility of the interaction between the electrode material and electrolyte are very crucial for achieving better performance of electrochemical supercapacitors. For example, Chmiola *et al.* and Largeot *et al.* observed that the proximity between average pore size of the carbon electrode and electrolyte ion size may enhance the capacitance of supercapacitors [88–91].

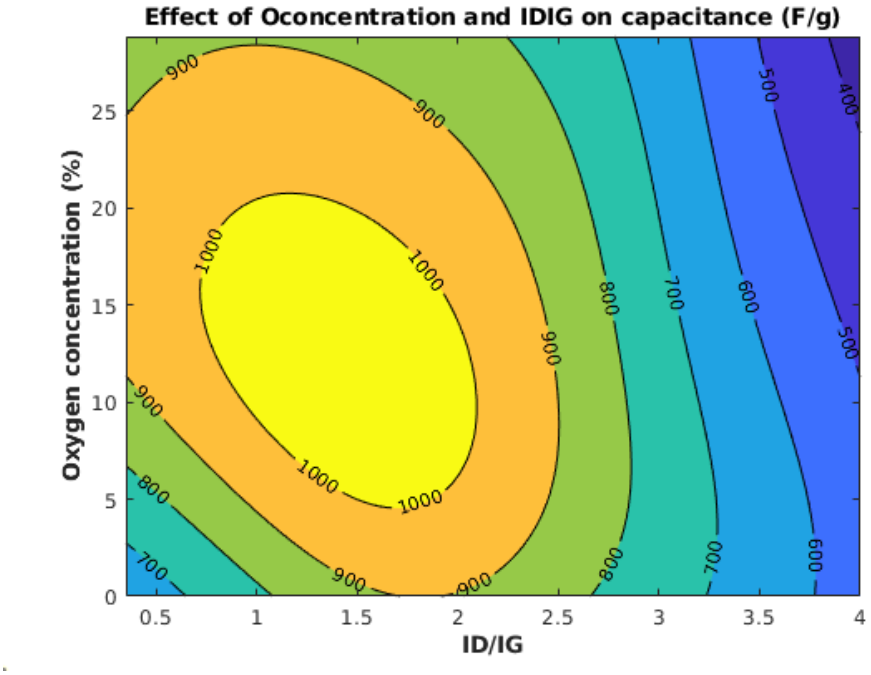

Figure 3.20. Effect of Oxygen concentration and ID/IG on capacitance.

Optimum electrolyte concentration is obtained as 4.8 M from ANN-GA hybrid system. The impact of the electrolyte concentration and SSA relationship on the capacitance is shown in the Figure 3.21. The figure implies that there is no best electrode and electrolyte material in supercapacitor devices. The design of optimal electrolyte/electrode material enables high performance supercapacitor. Also, it can be observed that optimum electrolyte concentration varies for different specific surface areas. Cheng *et al.* proposed using different electrolyte materials with the same electrode for graphene/CNTs composite. It is shown that the specific capacitance was obtained as 290 F/g in aqueous electrolyte and 201 F/g in organic electrolyte [92].

Also, utilizing electrolytes with wider operating voltage is another approach for boosting the energy density of supercapacitors. Ionic liquids are novel electrolyte materials since they provide a wider potential window than organic and aqueous electrolytes.The optimal solution of operating voltage is 4.7 V which is also reasonable since the increased voltage may affect the specific capacitance positively. Li *et al*

reported a high operating voltage (4.7 V) with ionic liquids for SWCNTs/graphene framework [93]. Therefore, optimum absolute potential window is also feasible for proposed supercapacitor design.

Current collector, separator, and binder concentration can be examined individually since they are not relevant to other parameters. Current collector should be light, non-reactive, and highly conductive to achieve high-performance supercapacitors. The optimum current collector of ANN-GA output is Au/Pt, and they are frequently utilized in carbon-based supercapacitors [9].

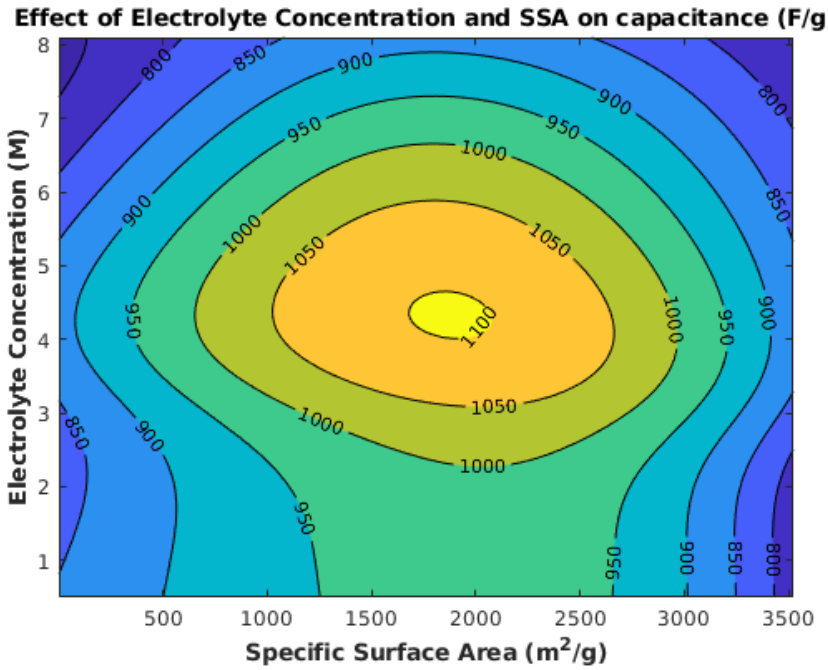

Figure 3.21. Effect of Electrolyte concentration and SSA on capacitance.

Separator material is Cellulose for the optimal solution. Separator should possess high porosity, electrical conductivity, and mechanical strength. Cellulose is one of the most efficient separators with respect to material selection analysis in supercapacitor applications. Also, it is combined with graphene/CNTs electrode material in previous studies [9,94]. It is worth noting that there is no restriction to combine optimal current collector and separator material. In general, carbon electrode is prepared by mixing binder material. PTFE is widely used as a binder for graphene/CNTs electrodes and

its amount 10% [72, 95]. The optimal solution is 8.4% that can be seen in Figure 3.22.

Moreover, genetic algorithms possess many advantages in terms of optimization techniques. GAs enable robust sampling for the model, seek an optimal solution with specific pattern and there is no need for derivative calculation. However, it should be considered that data-driven optimization might not give exact solution. Since the data is collected from experimental papers, human error should take into consideration in terms of collecting data and experimental errors. In addition, it is difficult to solve problems with integer decision variables and nonlinear equality constraints with MATLAB's "ga" function. Inequality constraints should represented as equality constraints. In some cases, genetic algorithms do not ensure the optimum solution, yet, they are able to find near-optimal simply and robustly. Therefore, they can be integrated as starting models for experimental or analytical methods as well.

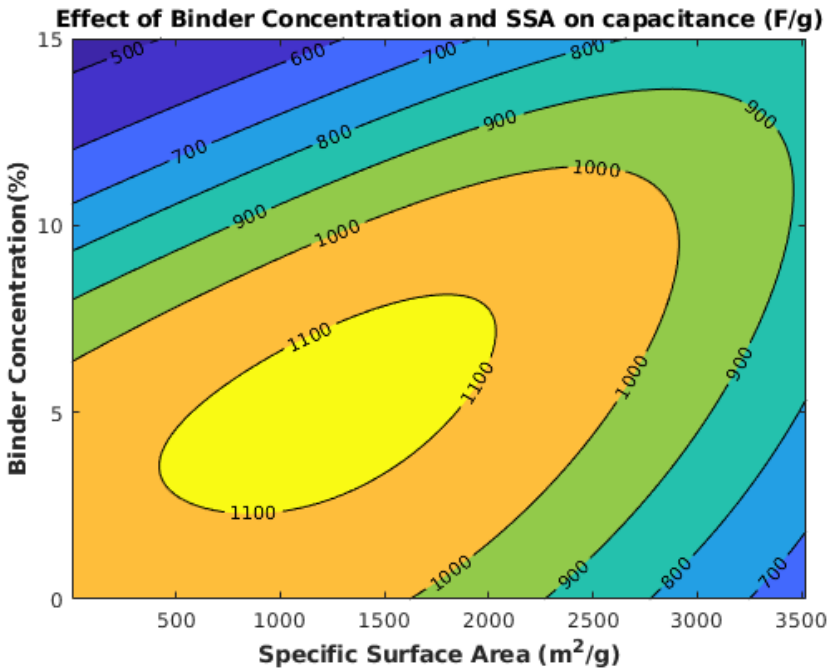

Figure 3.22. Effect of Binder Concentration and SSA on capacitance.

Also, experiments that would not possible under existing circumstances are conducted thanks to this model. Complex models may be simulated to capture and analyze difficult relationships [46]. Furthermore, the proposed ANN-GA hybrid model is a

generic algorithm and may utilize for different domains. Previous studies for ANN-GA algorithm do not use categorical features. Hence, this model can be applied as an optimization tool for mixed-integer nonlinear problems.

Up to now, the maximum capacitance value of graphene/CNTs-based supercapacitors is 557 F/g in the literature to the best of our knowledge. Sammed *et al.* reported reduced holey graphene oxide/CNT material via combination of chemical vapor depsition and vacuum filtration. The experiments are carried out in 6 M KOH aqueous electrolyte with potential window from -0.9 V to -0.1 V [96]. Hence, 1012 F/g can be achieved with suitable combinations of parameters by using nitrogen doping, wide operating voltage and high surface area. As a conclusion, the parameter set that is obtained from ANN/GA hybrid model is consistent with literature and optimal capacitance may be achieved by using proposed design of N-doped graphene/CNTs based supercapacitors.

4. CONCLUSION AND OUTLOOK

4.1. Conclusion

In recent years, supercapacitors have drawn great attention thanks to their high power density, long cycle life, and rapid charge/discharge rate. However, they suffer from low energy density. The primary focus regarding supercapacitors is enhancing the capacitance which leads to high energy density. The proposed ANN-GA is aimed to obtain a parameter set that yields optimum capacitance of supercapacitors.

For this purpose, a large amount of data was extracted from 110 published articles from the literature. Dataset is carefully examined and several data preprocessing techniques are implemented. Outlier detection is utilized for target and input variables. Min-max normalization is used in the range of [0, 1] to reduce bias. The dataset includes five categorical features that are described for supercapacitors such as electrode material, separator, current collector, etc. These categorical attributes are converted into numerical form by using one-hot encoding. Also, feature selection is performed to get rid of irrelevant features. Trial/error method, creating subset of features and random forest feature importance are employed as feature selection methods. Statistical learning approaches are computed to predict capacitance using selected features. Linear regression, lasso regression, regression tree, random forest, gradient boosting, and neural networks are applied. Performance metrics are defined to measure the performance of machine learning algorithms. Root mean square error and determination of coefficients are used as performance metrics. 5 and 10-fold cross-validation is introduced to evaluate consistency of models. Also, parameter tuning is carried out before the prediction process to reduce model complexity and prevent overfitting. Shrinkage parameter is tuned for lasso regression automatically. The minimal number of observations per tree leaf and the number of variables randomly sampled are tuned for regression tree and random forest, respectively. Grid search is used for gradient boosting and neural networks since they have two tuning parameters. Learning rate and

number of cycling are hyperparameters that affect the overall performance of gradient boosting. For neural networks, activation function and number of neurons are tuned to obtain robust model. All parameters are selected according to average cross-validation results that give minimum error.

Artificial neural network model with sigmoid activation function and 16 neurons demonstrated better fitting and lower error among attempted statistical learning approaches. It is expected that neural network models exhibit rigorous solutions for data sets that include highly correlated nonlinear relationships. Therefore, artificial neural networks model is used as the objective function of the genetic algorithm. Genetic algorithms are known as effective and practical search algorithms in which the computational procedure mimics the theory of biological evolution. Genetic algorithm is utilized to achieve optimum capacitance and feasible parameter set. Selection, crossover, and mutation operators are used to form new offspring and increase input space. The integration of a neural network model for prediction purposes with a genetic algorithm as an optimization tool possesses important potentials regarding several applications [49].

Genetic algorithm is computed with specific constraints that are defined by considering expert knowledge. Side constraints, linear equality /inequality constraint, and integer constraints for categorical variables. Random population is chosen from the supercapacitor dataset. Tournament selection, crossover, and mutation are used as main operators to get children solutions. Also, qualified solutions are kept in the population as an elite member generation by generation.

The candidate optimal solutions are observed via prespecified constraint by considering each electrode material. The optimum capacitance is achieved as 1012 F/g for N-doped graphene/CNT electrode material that satisfies specified constraints. Also, parameter set that gives optimum capacitance is obtained from ANN-GA hybrid model and feasibility of proposed solution is discussed. Literature information and design maps are used to assess the applicability of results. Design maps are created to examine effect of two variables on capacitance while other parameters are fixed to their

optimal values. It is deduced that optimum capacitance is accessible with the optimal values of input variables such as SSA, potential window, and nitrogen/oxygen doping.

As a conclusion, the hybrid ANN-GA model is an effective and practical tool to obtain high-performance supercapacitors. In addition to this, this model is generic and can be used in any different field. Artificial neural networks model successfully pursue nonlinear relationships of feature to predict output variable. Genetic algorithm provides an opportunity to evaluate novel ideas and combinations of various attributes. Also, ANN-GA based optimization algorithm offers alternative solution to a derivative-based optimization algorithm for nonlinear mixed integer problems.

4.2. Outlook

In this thesis, hybrid ANN-GA-based optimization model provides a practical and effective solution for the feasible design of supercapacitor that can achieve ultrahigh capacitance. However, the performance of the algorithm may improve by introducing several modifications. The proposed hybrid ANN-GA model might be extended by using a multi-objective genetic algorithm. For instance, the power density of the supercapacitor may be kept at the desired level or maximized while the capacitance is optimized. Also, sensitivity analysis can be carried out to determine the importance of inputs on capacitance. Sensitivity analysis (SA) allows us to gain deeper insight into the modeling by measuring the importance of each input variable on target. SA requires analysis and computation of sensitivity of input attributes by using given criteria on the response variable [50, 97]. In addition to this, the robustness of the model can be evaluated by using different methods to compare the results of the proposed model. Bayesian optimization is a powerful and sequential algorithm that benefits from previous steps for optimizing design problems. In particular, it is useful for design problems in which objective function is costly and nonconvex. BO framework comprises two main parameters that are Bayesian statistical model (usually Gaussian Process) and the acquisition function. Gaussian process is utilized to model objective function and acquisition function determines next sample [98,99]. BO is widely used in design prob-

lems such as fibril design and identification of cycle life for batteries [100, 101]. The proposed hybrid ANN-GA based optimization model offers a simple and derivative-free method to solve nonconvex mixed-integer problems. Despite the simplicity and applicability to many optimization problems, it requires specific implementations and adjustments according to the type of the problem [53]. For this reason, the reliable method is converting such a problem into a convex formulation that guarantees global optima. The convex formulation may provide a more general approach and avoid local optima.

REFERENCES

1. Vosen, S. and J. Keller, "Hybrid Energy Storage Systems for Stand-Alone Electric Power Systems: Optimization of System Performance and Cost Through Control", *International Journal of Hydrogen Energy*, Vol. 24, No. 12, pp. 1139–1156, 1999.

2. Choi, M.-E., S.-W. Kim and S.-W. Seo, "Energy Management Optimization in a Battery/Supercapacitor Hybrid Energy Storage System", *IEEE Transactions on Smart Grid*, Vol. 3, No. 1, pp. 463–472, 2012.

3. Ju, F., Q. Zhang, W. Deng and J. Li, "Review of Structures and Control of Battery-Supercapacitor Hybrid Energy Storage System for Electric Vehicles", *Advances in Battery Manufacturing, Services, and Management Systems*, pp. 303–318, 2016.

4. Vlad, A., N. Singh, J. Rolland, S. Melinte, P. M. Ajayan and J. F. Gohy, "Hybrid Supercapacitor-Battery Materials for Fast Electrochemical Charge Storage", *Scientific Reports*, Vol. 4, No. 4315, 2014.

5. Simon, P. and Y. Gogotsi, "Perspectives for Electrochemical Capacitors and Related Devices", *Nature Materials*, Vol. 19, No. 11, 2020.

6. Zhang, S. and N. Pan, "Supercapacitors Performance Evaluation", *Advanced Energy Materials*, Vol. 5, No. 6, p. 1401401, 2015.

7. Wayu, M., "Manganese Oxide Carbon-Based Nanocomposite in Energy Storage Applications", *Solids*, Vol. 2, No. 2, pp. 232–248, 2021.

8. Halper, M. S. and J. C. Ellenbogen, *Supercapacitors : A Brief Overview*, Mitre Corporation, 2006.

9. Kar, K., *Handbook of Nanocomposite Supercapacitor Materials I Characteristics: Characteristics*, Springer, 2020.

10. Simon, P. and Y. Gogotsi, "Materials for Electrochemical Capacitors", *Nature Materials*, Vol. 7, pp. 845–54, 2008.

11. Gheytanzadeh, M., A. Baghban, S. Habibzadeh, A. Mohaddespour and O. Abida, "Insights into the Estimation of Capacitance for Carbon-Based Supercapacitors", *RSC Advances*, Vol. 11, pp. 5479–5486, 2021.

12. Lee, W.-H. and J. H. Moon, "Monodispersed N-Doped Carbon Nanospheres for Supercapacitor Application", *ACS Applied Materials & Interfaces*, Vol. 6, pp. 13968–13976, 2014.

13. Ren, J., X. Lin, J. Liu, T. Han, C. Wang, H. Zhang and J. Li, "Engineering Early Prediction of Supercapacitors' Cycle Life using Neural Networks", *Materials Today Energy*, Vol. 18, p. 100537, 2020.

14. Burkov, A., *The Hundred-Page Machine Learning Book*, Andriy Burkov, 2019.

15. Zhu, S., J. Li, L. Ma, C. He, E. Liu, F. He, C. Shi and N. Zhao, "Artificial Neural Network Enabled Capacitance Prediction for Carbon-Based Supercapacitors", *Materials Letters*, Vol. 233, pp. 294–297, 2018.

16. Su, H., C. Lian, J. Liu and H. Liu, "Machine Learning Models for Solvent Effects on Electric Double Layer Capacitance", *Chemical Engineering Science*, Vol. 202, pp. 186–193, 2019.

17. Liu, P., Y. Wen, L. Huang, X. Zhu, R. Wu, S. Ai, T. Xue and Y. Ge, "An Emerging Machine Learning Strategy for the Assisted Design of High Performance Supercapacitor Materials by Mining the Relationship between Capacitance and Structural Features of Porous Carbon", *Journal of Electroanalytical Chemistry*, Vol. 899, pp. 115–684, 2021.

18. Ahmed, S., S. Radhakrishnan, B. Nair and R. Thiruvengadathan, "Efficient Hyperparameter-Tuned Machine Learning Approach for Estimation of Supercapacitor Performance Attributes", *Journal of Physics Communications*, Vol. 5, 2021.

19. Farsi, H. and F. Gobal, "Artificial Neural Network Simulator for Supercapacitor Performance Prediction", *Computational Materials Science*, Vol. 39, p. 678–683, 2007.

20. Ding, S., J. Hang, B. Wei and Q. Wang, "Modelling of Supercapacitors Based on SVM and PSO Algorithms", *IET Electric Power Applications*, Vol. 12, pp. 502–507, 2018.

21. Mirjalili, S., "Genetic Algorithm", *Studies in Computational Intelligence*, pp. 43–55, 2019.

22. Zhou, Y., Y. Wang, L. Kang, F. Peng, L. Wang and J. Pang, "Hybrid Genetic Algorithm Method for Efficient and Robust Evaluation of Remaining Useful Life of Supercapacitors", *Applied Energy*, Vol. 260, pp. 114–169, 2020.

23. Ghosh, S., R. R. Gangavarapu and T. Thomas, "Machine Learning-Based Prediction of Supercapacitor Performance for a Novel Electrode Material: Cerium Oxynitride", *Energy Storage Materials*, Vol. 40, pp. 426–438, 2021.

24. González, A., E. Goikolea, J. A. Barrena and R. Mysyk, "Review on Supercapacitors: Technologies and Materials", *Renewable and Sustainable Energy Reviews*, Vol. 58, pp. 1189–1206, 2016.

25. Famili, A., W.-M. Shen, R. Weber and E. Simoudis, "Data Preprocessing and Intelligent Data Analysis", *Intelligent Data Analysis*, Vol. 1, No. 1, pp. 3–23, 1997.

26. García, S., J. Luengo and F. Herrera, *Data Preprocessing in Data Mining*, Vol. 72,

Springer, 2015.

27. Han, J., M. Kamber and J. Pei, *Data Preprocessing*, pp. 83–124, Elsevier Science, 2012.

28. James, G., D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, Springer Series in Statistics, Springer New York, 2014.

29. Li, J., K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang and H. Liu, "Feature Selection: A Data Perspective", *ACM Computing Surveys*, Vol. 50, No. 6, pp. 1–45, 2017.

30. Vishwakarma, G., A. Sonpal and J. Hachmann, "Metrics for Benchmarking and Uncertainty Quantification: Quality, Applicability, and Best Practices for Machine Learning in Chemistry", *Trends in Chemistry*, Vol. 3, No. 2, pp. 146–156, 2021.

31. Hastie, T., R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer, 2009.

32. Cutler, A., D. Cutler and J. Stevens, *Tree-Based Methods*, pp. 1–19, Springer Science, 2008.

33. Kern, C., T. Klausch and F. Kreuter, "Tree-Based Machine Learning Methods for Survey Research", *Survey Research Methods*, Vol. 13, pp. 73–93, 2019.

34. Breiman, L., "Random Forests", *Machine Learning*, Vol. 45, pp. 5–32, 2001.

35. Friedman, J., "Greedy Function Approximation: A Gradient Boosting Machine", *The Annals of Statistics*, Vol. 29, No. 5, p. 1189–1232, 2001.

36. Burke, L. I., "Introduction to Artificial Neural Systems for Pattern Recognition", *Computers & Operations Research*, Vol. 18, No. 2, pp. 211–220, 1991.

37. Chow, T., G. Zhang, Z. Lin and C. Song, "Global Optimization of Absorption Chiller System by Genetic Algorithm and Neural Network", *Energy and Buildings*, Vol. 34, No. 1, pp. 103–109, 2002.

38. Goodfellow, I., Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

39. Bishop, C. M., *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006.

40. Reed1999, R. D. and R. J. Marks, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*, MIT Press, 1999.

41. Liu, H., "On the Levenberg-Marquardt Training Method for Feed-Forward Neural Networks", *2010 Sixth International Conference on Natural Computation*, Vol. 1, pp. 456–460, 2010.

42. Yu, H. and B. Wilamowski, *Levenberg–Marquardt Training*, pp. 1–16, CRC Press, 2011.

43. Goldberg, D., G. David Edward, D. Goldberg and V. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Artificial Intelligence, Addison-Wesley Publishing Company, 1989.

44. Mitchell, M., *An Introduction to Genetic Algorithms*, Complex Adaptive Systems, MIT Press, 1998.

45. Miller, B. L., B. L. Miller, D. E. Goldberg and D. E. Goldberg, "Genetic Algorithms, Tournament Selection, and the Effects of Noise", *Complex Systems*, Vol. 9, pp. 193–212, 1995.

46. Gallagher, K. and M. Sambridge, "Genetic Algorithms: A Powerful Tool for Large-Scale Nonlinear Optimization Problems", *Computers & Geosciences*, Vol. 20, No. 7, pp. 1229–1236, 1994.

47. Deep, K., K. Singh, M. Kansal and C. Mohan, "A Real Coded Genetic Algorithm for Solving Integer and Mixed Integer Optimization Problems", *Applied Mathematics and Computation*, Vol. 212, pp. 505–518, 2009.

48. Ghaedi, M., N. Zeinali, A. Ghaedi, M. Teimuori and J. Tashkhourian, "Artificial Neural Network-Genetic Algorithm Based Optimization for the Adsorption of Methylene Blue and Brilliant Green from Aqueous Solution by Graphite Oxide Nanoparticle", *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, Vol. 125, pp. 264–277, 2014.

49. Cook, D., C. Ragsdale and R. Major, "Combining a Neural Network with a Genetic Algorithm for Process Parameter Optimization", *Engineering Applications of Artificial Intelligence*, Vol. 13, No. 4, pp. 391–396, 2000.

50. Salehi, M., S. Farhadi, A. Moieni, N. Safaie and H. Ahmadi, "Mathematical Modeling of Growth and Paclitaxel Biosynthesis in Corylus avellana Cell Culture Responding to Fungal Elicitors Using Multilayer Perceptron Genetic Algorithm", *Frontiers in Plant Science*, Vol. 11, 2020.

51. Biegler, L. T. and I. E. Grossmann, "Retrospective on Optimization", *Computers & Chemical Engineering*, Vol. 28, No. 8, pp. 1169–1192, 2004.

52. Grossmann, I., "Review of Non-Linear Mixed Integer and Disjunctive Programming Techniques for Process Systems Engineering", *Optimization and Engineering*, Vol. 3, pp. 227–252, 2002.

53. Belotti, P., C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke and A. Mahajan, "Mixed-integer Nonlinear Optimization", *Acta Numerica*, Vol. 22, p. 1–131, 2013.

54. Gantovnik, V., Z. Gurdal, L. Watson and C. Anderson-Cook, "A Genetic Algorithm for Mixed Integer Nonlinear Programming Problems Using Separate Constraint Approximations", *AIAA Journal*, Vol. 43, pp. 1844–1849, 2003.

55. Lee, J. and S. Leyffer, *Mixed Integer Nonlinear Programming*, The IMA Volumes in Mathematics and its Applications, Springer New York, 2011.

56. Belotti, P., P. Bonami, M. Fischetti, A. Lodi, M. Monaci, A. Nogales-Gómez and D. Salvagnin, "On Handling Indicator Constraints in Mixed Integer Programming", *Computational Optimization and Applications*, Vol. 65, pp. 545–566, 2016.

57. Bonami, P., A. Lodi, A. Tramontani and S. Wiese, "On Mathematical Programming with Indicator Constraints", *Mathematical Programming*, Vol. 151, pp. 191–293, 2015.

58. Griva, I., S. Nash and A. Sofer, *Linear and Nonlinear Optimization: Second Edition*, Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, 2009.

59. Saadetnejad, D., D. T. Karabay and B. Uralcan, "Machine Learning Modelling of EDLCs", *Manuscript in Preparation*, 2022.

60. Hastie, T., R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2001.

61. Nascimento, C. A. O., R. Giudici and R. Guardani, "Neural Network Based Approach for Optimization of Industrial Chemical Processes", *Computers & Chemical Engineering*, Vol. 24, No. 9, pp. 2303–2314, 2000.

62. Rogers, J. and S. Gunn, "Identifying Feature Relevance Using a Random Forest", *Subspace, Latent Structure and Feature Selection*, pp. 173–184, 2005.

63. Jha, N., P. Ramesh, E. Barkyarova, M. Itkis and R. Haddon, "High Energy Density Supercapacitor Based on a Hybrid Carbon Nanotube-Reduced Graphite Oxide Architecture", *Advanced Energy Materials*, pp. 438–444, 2012.

64. Yang, Z., J. Tian, Z. Yin, C. Cui, W. Qian and F. Wei, "Carbon Nanotube- and Graphene-Based Nanomaterials and Applications in High-Voltage Supercapacitor: A Review", *Carbon*, Vol. 141, pp. 467–480, 2019.

65. Zhu, Y., L. Li, C. Zhang, G. Casillas, Z. Sun, Z. Yan, G. Ruan, Z. Peng, A.-R. Raji, C. Kittrell, R. Hauge and J. Tour, "A Seamless Three-Dimensional Carbon Nanotube Graphene Hybrid Material", *Nature Communications*, Vol. 3, p. 1225, 2012.

66. Bi, Z., Q.-Q. Kong, Y. Cao, G. Sun, F. Su, X.-X. Wei, X. li, A. Ahmad, L. Xie and C.-M. Chen, "Biomass-Derived Porous Carbon Materials with Different Dimensions for Supercapacitor Electrodes: A Review", *Journal of Materials Chemistry A*, 2019.

67. Su, F., A. Chee, C. Poh, G. Xu, D. Wang, Q. Li, J. Lin, J. S. Chen and X. Lou, "Nitrogen-Containing Microporous Carbon Nanospheres with Improved Capacitive Properties", *Energy & Environment Sciences*, Vol. 4, pp. 717–724, 2011.

68. Peng, H., G. Ma, K. Sun, J. Mu, X. Zhou and Z. Lei, "A Novel Fabrication of Nitrogen-Containing Carbon Nanospheres with High Rate Capability as Electrode Materials for Supercapacitors", *RSC Advances*, Vol. 5, pp. 12034–12042, 2015.

69. Ruiz, V., C. Blanco, M. Granda, R. Menéndez and R. Santamaría, "Effect of the thermal treatment of carbon-based electrodes on the electrochemical performance of supercapacitors", *Journal of Electroanalytical Chemistry*, Vol. 618, No. 1, pp. 17–23, 2008.

70. Picó, F., J. Rojo, M. Sanjuán, A. Ansón, A. Benito, M. Callejas, W. Maser and

M. Martínez, "Single-Walled Carbon Nanotubes as Electrodes in Supercapacitors", *Journal of the Electrochemical Society*, Vol. 151, 2004.

71. Wang, Z., H. Qiang, Z.-h. Zhu, J. Liu, C. Chen and D. Zhang, "Facile Synthesis of Nitrogen-Doped Mesoporous Hollow Carbon Nanospheres for High-Performance Supercapacitors", *ChemElectroChem*, Vol. 5, pp. 2242–2249, 2018.

72. Ye, J., B. Hu, Y. Jin, Z. Wang, Y. Xi, F. Liang and Q. Pan, "Interface Engineering Integrates Fractal-Tree Structured Nitrogen-Doped Graphene/Carbon Nanotubes for Supercapacitors", *Electrochimica Acta*, Vol. 349, p. 136372, 2020.

73. You, B., L. Wang, L. Yao and J. Yang, "Three Dimensional N-Doped Graphene-CNT Networks for Supercapacitor", *Chemical Communications (Cambridge, England)*, Vol. 49, 2013.

74. Ye, Z., F. Wang, C. Jia, K. Mu, M. Yu, Y. Lv and Z. Shao, "Nitrogen and Oxygen-Codoped Carbon Nanospheres for Excellent Specific Capacitance and Cyclic Stability Supercapacitor Electrodes", *Chemical Engineering Journal*, Vol. 330, p. 1166–1173, 2017.

75. Li, M., Y. Zhang, L. Yang, Y. Liu and J. Yao, "Hollow Melamine Resin-Based Carbon Spheres/Graphene Composite with Excellent Performance for Supercapacitors", *Electrochimica Acta*, Vol. 166, pp. 310–319, 2015.

76. Sun, H.-J., B. Liu, T.-J. Peng and X.-L. Zhao, "Nitrogen-Doped Porous 3D Graphene with Enhanced Supercapacitor Properties", *Journal of Materials Science*, Vol. 53, pp. 13100–13110, 2018.

77. Li, Y., S. Zhang, H. Song, X. Chen, J. Zhou and S. Hong, "New Insight into the Heteroatom-Doped Carbon as the Electrode Material for Supercapacitors", *Electrochimica Acta*, Vol. 180, pp. 879–886, 2015.

78. Ilnicka, A., M. Skorupska, M. Szkoda, Z. Zarach, P. Kamedulski, W. Zieliński

and J. Lukaszewicz, "Combined Effect of Nitrogen-Doped Functional Groups and Porosity of Porous Carbons on Electrochemical Performance of Supercapacitors", *Scientific Reports*, Vol. 11, p. 18387, 2021.

79. Yang, M. and Z. Zhou, "Recent Breakthroughs in Supercapacitors Boosted by Nitrogen-Rich Porous Carbon Materials", *Advanced Science*, Vol. 4, No. 8, p. 1600408, 2017.

80. Jeong, H., J. Lee, W. Shin, Y. Choi, H.-J. Shin, J. Kang and J. Choi, "Nitrogen-Doped Graphene for High-Performance Ultracapacitors and the Importance of Nitrogen-Doped Sites at Basal Planes", *Nano Letters*, Vol. 11, pp. 2472–2477, 2011.

81. Paek, E., A. Pak, K. Kweon and G. Hwang, "On the Origin of the Enhanced Supercapacitor Performance of Nitrogen-Doped Graphene", *The Journal of Physical Chemistry C*, Vol. 117, pp. 5610–5616, 2013.

82. Jonna, N., M. Janakiraman, S. S. Saravananbhavan, K. Natarajan, P. Cao and B. Natesan, "Nitrogen-Doped Graphene/Poly(3,4-ethylenedioxythiophene) as Metal-Free Electrode Material for High-Performance Supercapacitor Applications", *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1–14, 2020.

83. Sun, L., L. Wang, C. Tian, T. Tan, Y. Xie, K. Shi, M. Li and H. Fu, "Nitrogen-Doped Graphene with High Nitrogen Level Via a One-Step Hydrothermal Reaction of Graphene Oxide with Urea for Superior Capacitive Energy Storage", *RSC Advances*, Vol. 2, pp. 4498–4506, 2012.

84. Lin, T.-T., W.-H. Lai, Q.-F. Lü and Y. Yu, "Porous Nitrogen-Doped Graphene/Carbon Nanotubes Composite with an Enhanced Supercapacitor Performance", *Electrochimica Acta*, Vol. 178, pp. 517–524, 2015.

85. Punckt, C., M. Pope, W. Liu, Y. Lin and I. Aksay, "Electrochemical Performance of Graphene as Effected by Electrode Porosity and Graphene Functionalization", *Electroanalysis*, Vol. 22, pp. 2834 – 2841, 2010.

86. Cao, H., X. Peng, M. Zhao, P. Liu, B. Xu and J. Guo, "Oxygen Functional Groups Improve the Energy Storage Performances of Graphene Electrochemical Supercapacitors", *RSC Advances*, Vol. 8, pp. 2858–2865, 2018.

87. Pope, M., C. Punckt and I. Aksay, "Intrinsic Capacitance and Redox Activity of Functionalized Graphene Sheets", *The Journal of Physical Chemistry C*, Vol. 115, pp. 20326–20334, 2011.

88. Pal, B., S. Yang, S. Ramesh, V. Thangadurai and R. Jose, "Electrolyte Selection for Supercapacitive Devices: A Critical Review", *Nanoscale Advances*, Vol. 1, 2019.

89. Zhang, J. and X. Zhao, "On the Configuration of Supercapacitors for Maximizing Electrochemical Performance", *ChemSusChem*, Vol. 5, pp. 818–41, 2012.

90. Chmiola, J., G. Yushin, R. Dash and Y. Gogotsi, "Effect of Pore Size and Surface Area of Carbide Derived Carbons on Specific Capacitance", *Journal of Power Sources*, Vol. 158, pp. 765–772, 2006.

91. Largeot, C., C. Portet, J. Chmiola, P.-L. Taberna, Y. Gogotsi and P. Simon, "Relation between the Ion Size and Pore Size for an Electric Double-Layer Capacitor", *Journal of the American Chemical Society*, Vol. 130, pp. 2730–2731, 2008.

92. Cheng, Q., J. Tang, J. Ma, H. Zhang, N. Shinya and L.-C. Qin, "Graphene and Carbon Nanotube Composite Electrodes for Supercapacitors with Ultra-High Energy Density", *Physical Chemistry Chemical Physics*, Vol. 13, pp. 17615–17624, 2011.

93. Li, J., J. Tang, J. Yuan, K. Zhang, Y. Sun, H. Zhang and L.-C. Qin, "Enlarging

Energy Density of Supercapacitors using Unequal Graphene Electrodes and Ionic Liquid Electrolyte", *Electrochimica Acta*, Vol. 258, pp. 1053–1058, 2017.

94. Borges, R., H. Ribeiro, R. Lavall and G. Silva, "Temperature Stable Supercapacitors Based on Ionic Liquid and Mixed Functionalized Carbon Nanomaterials", *Journal of Solid State Electrochemistry*, Vol. 16, pp. 3573–3580, 2012.

95. Ansaldo, A., P. Bondavalli, S. Bellani, A. Del Rio Castillo, M. Prato, V. Pellegrini, G. Pognon and F. Bonaccorso, "High-Power Graphene-Carbon Nanotube Hybrid Supercapacitors", *ChemNanoMat*, Vol. 3, pp. 436–446, 2017.

96. Sammed, K. A., L. Pan, M. Asif, M. Usman, T. Cong, F. Amjad and M. Imran, "Reduced Holey Graphene Oxide Film and Carbon Nanotubes Sandwich Structure as a Binder-Free Electrode Material for Supercapcitor", *Scientific Reports*, Vol. 10, 2020.

97. Iooss, B. and A. Saltelli, *Introduction to Sensitivity Analysis*, pp. 1–20, Springer International Publishing, 2015.

98. Shahriari, B., K. Swersky, Z. Wang, R. P. Adams and N. de Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization", *Proceedings of the IEEE*, Vol. 104, No. 1, pp. 148–175, 2016.

99. Frazier, P., "A Tutorial on Bayesian Optimization", *ArXiv*, Vol. abs/1807.02811, 2018.

100. Severson, K. A., P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh and R. D. Braatz, "Data-Driven Prediction of Battery Cycle Life before Capacity Degradation", *Nature Energy*, Vol. 4, No. 5, 2019.

101. Son, D., V. Liimatainen and M. Sitti, "Machine Learning-Based and Experimentally Validated Optimal Adhesive Fibril Designs", *Small*, Vol. 17, No. 39, p.

2102867, 2021.