Master thesis

Long-term Forecasting Heat Use in Sweden's Residential Sector using Genetic Algorithms and Neural Network

Authors: Mohammad Befkin & Alireza Momtaz
Supervisor: Mohammad Saeid Atabaki
Examiner: Mohsen Soleimani Mohseni
Abstract

In this study the parameters of population, gross domestic product (GDP), heat price, U-value, and temperature have been used to predict heat consumption for Sweden till 2050. It should be noted that the heat consumption has been considered for multi-family house. Most multi-family houses (MFH) get their primary heat from district heating (DH). A literature analysis of various models and variables has been conducted to enhance comprehension of forecasting and its process. The majority of earlier research has focused on electricity or energy rather than heat. The aim of this study is to create a model (linear and non-linear) from 1993 to 2019 with a minimum error as possible, and then use the genetic algorithm (GA) and neural network (NN) to predict Sweden's heat consumption till 2050.

Keywords: Genetic Algorithm; Neural Network; Forecasting; Heat use
Summary

Energy consumption modeling is based on past consumption and energy consumption is dependent on variables such as economic variables, demographic variables, weather indicators, etc. (Egelioglu et al, 2001). Currently, energy modeling is a broad topic that is of interest to many scientists and engineers, which leads them to pay attention to the problem of energy production and consumption (Ozturk et al, 2006). Modeling is useful and efficient in some fields of application for formulating policies and plans (Dincer & Dost, 1996). In this regard, energy planning is not possible without having an acceptable knowledge of past and present energy consumption and possible future demand (Sozen & Arcaklioglu, 2007). This article uses a neural network and genetic algorithm to forecast Sweden's future heat consumption. For this reason, the authors have compiled a variety of data from various years, and they have created a model that will be used for forecasting. Following data gathering, the authors started the processing procedure, which included training, testing, and normalizing. The model has been created using neural networks and genetic algorithms. After the model was created, various errors were utilized for evaluation, including mean error (ME), root mean square error (RMSE) and relative standard error (RSE) and the methods employed were compared with each other.

Keywords: Energy Modelling; Heat Consumption; Neural Network; testing; training
Preface

This master's thesis represents our earnest efforts and dedication to understanding complex topics like Genetic Algorithms and Neural Networks. We are deeply grateful to our supervisor, Dr. Mohammad Saeid Atabaki, for his unwavering guidance, expertise, and patience throughout this journey. We also express our appreciation to our esteemed professors and mentors for their insights and suggestions. Collaborating as co-authors has been an enlightening experience, and we aspire for this work to contribute meaningfully to the academic community.

Halmstad, Nov 2023
Mohammad Befkin & Alireza Momtaz
# Table of contents

1. Introduction .................................................................................................................. 7  
   1.1. Definition of the problem ...................................................................................... 8  
   1.2. Research questions ............................................................................................. 9  
   1.3. Research approach ............................................................................................. 9  
      1.3.1. Desk Research ......................................................................................... 9  
      1.3.2. Case Study .............................................................................................. 9  
2. Literature Review ........................................................................................................ 10  
3. Method of research ..................................................................................................... 18  
   3.1. Step 1: Collecting data ....................................................................................... 18  
   3.2. Step 2: Pre-Processing ...................................................................................... 19  
   3.3. Step 3: Training and testing .............................................................................. 19  
   3.4 Step 4: Modelling and forecasting ..................................................................... 20  
   3.5 Step 5: Models Evaluation ............................................................................... 20  
4. Method of Analysis ..................................................................................................... 21  
   4.1. Artificial neural network ................................................................................... 21  
   4.2 Details of Neural Network ............................................................................... 22  
   4.3 Optimization by Genetic Algorithm .................................................................. 29  
      4.3.1 Implementation ......................................................................................... 29  
5. Combined model of genetic algorithm and artificial neural network .................. 32  
6. Analysis and Results .................................................................................................. 35  
   6.1. Data Collection ................................................................................................. 35  
   6.2. Results ............................................................................................................. 39  
7. Innovation of the Thesis ............................................................................................. 46  
8. Conclusion .................................................................................................................. 51  
References ....................................................................................................................... I  
Appendix .......................................................................................................................... VIII
List of abbreviations in thesis

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF</td>
<td>Activation Function</td>
</tr>
<tr>
<td>AI</td>
<td>Artificial Intelligence</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>BNN</td>
<td>Biological Neural Networks</td>
</tr>
<tr>
<td>DH</td>
<td>District Heating</td>
</tr>
<tr>
<td>FFNN</td>
<td>Feed Forward Neural Network</td>
</tr>
<tr>
<td>FF</td>
<td>Fitness Function</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>IEA</td>
<td>International Energy Agency</td>
</tr>
<tr>
<td>LRM</td>
<td>Linear Regression Model</td>
</tr>
<tr>
<td>LSTM</td>
<td>Long-Term Short-Term Memory Model</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>ME</td>
<td>Mean Error</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>MFH</td>
<td>Multi-Family House</td>
</tr>
<tr>
<td>MLRM</td>
<td>Multivariate Linear Regression Model</td>
</tr>
<tr>
<td>MR</td>
<td>Multivariate Regression</td>
</tr>
<tr>
<td>MRA</td>
<td>Multivariate Regression Analysis</td>
</tr>
<tr>
<td>MRM</td>
<td>Multivariate Regression Model</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>NNM</td>
<td>Neural Network Model</td>
</tr>
<tr>
<td>SFH</td>
<td>Single-Family House</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>RNN</td>
<td>Recurrent Neural Network</td>
</tr>
<tr>
<td>RM</td>
<td>Regression Model</td>
</tr>
<tr>
<td>RI</td>
<td>Relative Improvement</td>
</tr>
<tr>
<td>RSE</td>
<td>Relative Standard Error</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
</tbody>
</table>
I. Introduction
The twentieth century observed a shift from resources based on coal to those based on petroleum (Suganthi & Samuel., 2012). Before 2030, there will be a more than 50% rise in global energy consumption if the current trend of consumption persists (Smith et al., 2007). Even if this prediction seems dire, it is still helpful to foresee the future since it might offer decision-makers crucial information because it has the potential to alter the current course that might result in disaster and catastrophe. A fundamental strategy for reducing the environmental effect of the construction industry, which accounts for 40% of society's energy consumption, is to reduce energy use in buildings (Wall, M., 2006). Heat and electricity are the natural topics of conversation when discussing energy. This study proposes a novel method for predicting the demand for heat. There are many ways to lessen the need for heat in Sweden, including improving building sector efficiency, lowering U-values, governmental policies, etc. U-value is defined as the amount of heat that passes through one square meter of a structure when the temperatures on each side of it differ by one degree Celsius (Baker, P., 2011). Its also known as the thermal transmittance coefficient (Baker, P., 2011). Moreover, with the aid of accurate predictions, decision-makers may better organize and plan the operations of the supply system by understanding the volume and trend of future demand (Ghalekhondabi et al., 2017). As Sweden moves towards renewable energy and net zero-emission buildings, accurate heat consumption forecasting becomes increasingly important. Still, this transition requires government and municipalities to take the lead and offer a pathway for both suppliers and customers and encourage them by using various strategies like tax reduction, subsidies, etc. Municipalities have a critical and essential role in Sweden. They serve as a bridge between decision-makers and the general public as well as the business community. The municipality executes decisions made by decision-makers, blends those decisions with the local context, and converts regional and national aims into local objectives and measures (Andersson, et al., 2019). There are several forecasting techniques, including Particle Swarm Optimization, Neural Networks (NN), and Genetic algorithm (GA) which are frequently used for time series forecasting. Each technique has advantages and drawbacks. In order to forecast Sweden's future heating demands, the authors have opted to focus on the GA model due to its adaptability in optimizing complex problems, capability to explore a vast solution space, and its robustness against local optima, making it suitable for the intricacies involved in heat demand predictions. GA is an algorithm that uses population data. Each parameter represents a gene, and every solution corresponds to a chromosome. Using a fitness function (FF), GA assesses the fitness of each member of the population (Mirjalili, S. and Mirjalili, S., 2019). The three stages of the GA procedure are selection, crossover, and mutation. The optimum solution was discovered during an iterative process using MATLAB software. Population,
GDP, U-value, and temperature are used in this research to predict heat consumption for Sweden till 2050. The focus area is heat consumption for multi-family buildings (MFB).

### 1.1. Definition of the problem

The increase in global demand for energy requires the development of intelligent forecasting methods and algorithms. Estimation of energy demand is based on economic and non-economic indicators that may be obtained using linear and non-linear statistical methods, mathematics and simulation models. Due to the nonlinearity of these indicators and energy consumption, researchers have turned to nonlinear modeling and prediction techniques including fuzzy regression, NN, and GA for solutions (Azadeh et al., 2008). Energy consumption modeling is usually based on past consumption and energy consumption is dependent on variables such as economic variables, demographic variables, weather indicators, etc. (Egelioglu et al, 2001). Currently, energy modeling is a broad topic that is of interest to many scientists and engineers, which leads them to pay attention to the problem of energy production and consumption (Ozturk et al, 2006). Modeling is useful and efficient in some fields of application for formulating policies and plans (Dincer & Dost, 1996). In this regard, energy planning is not possible without having an acceptable knowledge of past and present energy consumption and possible future demand (Sozen & Arcaklioglu, 2007). Energy consumption modeling and forecasting plays a very important role in the development and progress of countries for policy makers and related organizations. Ignoring consumption leads to potential energy outages that destroy environment and the economy. Overestimating energy may lead to wasteful and unnecessary capacity building, which means wasted financial resources. Therefore, it is better to use models that estimate energy consumption with higher accuracy in order to avoid costly mistakes. It is also better to use models that can use energy consumption data that are non-linear in nature in forecasting (Kavaklioglu et al, 2009). According to past studies, regression analysis has been considered as the most popular technique in predicting energy consumption, but the approach of artificial neural networks (ANN) is more attractive and important for potential users such as energy engineers. The reason is that regardless of the reduction of the required time, it provides the possibility of creating more stable energy applications. Also, this approach has advantages such as fast calculation, low cost and easy design by operators with little technical experience. Therefore, the use of ANN for modeling and forecasting is a goal that has been increasingly noticed in the last decade. It can be said that this is mainly because ANN has more advantages such as fast processing time and short development and very good capabilities for estimation. Especially, ANN is very useful in predicting problems that are unknown and mathematical formulas and previous knowledge about its inputs and outputs do not exist (Sozen & Arcaklioglu, 2007; Murat & Ceylan, 2006;
Sozen et al., 2005). This thesis focuses on modeling and forecasting energy consumption in Sweden based on energy, economic, and demographic indicators. Different models are studied in order to determine the best possible method in modeling energy consumption as a function of different indicators. Then, by using the methods of ANN, multivariate regression (MR) and GA, the energy consumption of Sweden is predicted so that finally the effectiveness of these methods in predicting energy consumption can be compared. The results of this thesis provide the main and important materials in the evaluation of energy consumption patterns for profitable companies and also provide the possibility of choosing a more accurate approach to estimate energy consumption in the future.

1.2. Research questions
1) What are the variables affecting heat consumption in Sweden?
2) What is the appropriate ANN model to predict Sweden's heat consumption?
3) Which of the regression analysis models are more effective in predicting heat consumption?

1.3. Research approach
The research approach adopted in this thesis encompasses desk research and case study approaches.

1.3.1. Desk Research
This method was selected because the factors affecting energy consumption were synthesized from existing research literature.

1.3.2. Case Study
The Swedish heat supply is analyzed using various statistical sources, which aids in understanding the unique context of energy consumption in Sweden. Data was primarily derived from the Ministry of Energy and the Central Bank. Additional databases and sources of data supply include:

- A comprehensive review of the relevant literature
- The International Energy Agency (IEA) databases
- Studying the documents of the Ministry of Energy and the Central Bank

The research process initiated with the design of a model based on studies in the domain of forecasting energy consumption and influential variables. Subsequently, energy consumption predictions were made using regression analysis and ANN. These methods were evaluated for efficiency and subsequently compared. In the final phase, a hybrid model of GA and NN was employed for forecasting, and its efficiency was juxtaposed against the prior two methods.
2. Literature Review
The subject of the causal relationship between energy consumption and population indicators has been well studied in the literature of energy economics (Sözen & Arcaklioglu, 2006). Indicators used in the previous studies has shown in Fig. 2.1.

Modelling along with co-accumulation technique or multivariate regression analysis (MRA) has been used in a large number of studies conducted in different countries to examine the effect of various factors on energy consumption. Also, various methods were used to predict energy consumption, some of these studies are as follows:

Glasure & Lee (1997) investigated the two-way causal relationship between energy consumption and GDP using co-integration and error models. They conducted their research in South Korea and Singapore. The results show no causal relationship between GDP and energy consumption for South Korea and unidirectional causal relationship from energy consumption to GDP for Singapore (Glasure & Lee, 1997). They concluded that The Granger causality test between GDP and energy consumption can be expressed as the following equation.

\[
\Delta y = \alpha + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \sum_{i=1}^{k} \gamma_i \Delta e_{t-i} + \mu_i
\]

Eq. 2.1

In which:

\( y \) and \( e \) are GDP and energy consumption. \( \mu \) is a disturbance term and other parameters are the coefficients to be estimated.

Ranjan & Jain (1999) describe energy consumption patterns for New Delhi, India as a function of climate and population. The period 1984–1993 has been analyzed by them using different linear regression models (LRM). The models were considered variations of consumption by the seasons (Ranjan & Jain, 1999). The following equation were obtained based on their research results:

![Fig. 2.1: Indicators used in the previous studies (Sözen & Arcaklioglu, 2006)](image-url)
Winter: \[ Y = -25.468 + 5.964 \times (POP) - 0.338 \times (T) \]

Summer: \[ Y = -44.219 + 6.992 \times (POP) + 0.244 \times (T) \] \hspace{1cm} \text{Eq. 2.2}

In which:
Y is the electrical energy consumption; monthly average values of population represent as (POP), temperature is (T).

Yang (2000) in his study in Taiwan, found a two-way causal relationship between energy consumption and GDP. He studied relation between energy consumption including coal, oil, natural gas, and electricity by GDP using Taiwan data for the period 1954–1997. Results showed bi-directional causality between total energy consumption and GDP (Yang, 2000).

Soytas & Sari (2003) examined the relationship between energy consumption and GDP in more than 11 emerging markets and G-7 countries excluding China due to lack of data. They discovered bi-directional relation in Argentina, relation from GDP to energy consumption in Italy and Korea and from energy consumption to GDP in Turkey, France, Germany and Japan (Soytas & Sari, 2003).

Narayan & Smyth (2005) investigated the relationship between electricity consumption, employment, and real income in Australia in the framework of causality and co-accumulation. Results showed that these parameters are co-integrated in the long run but there is weak unidirectional relation in the short run (Narayan & Smyth, 2005). The actual, fitted and residuals of electricity consumption have been shown in Fig. 2.2.

![Fig. 2.2: The actual, fitted and residuals of electricity consumption (Narayan & Smyth, 2005)](image)

Using data from 1980 to 2005, Jinke et al. (2008) investigated the differences in the cause-and-effect relationship between coal consumption and GDP in most OECD member and non-member countries as shown in Fig. 2.3. Results showed that unidirectional relation running from GDP to coal consumption
in Japan and China, and no relationship in India, South Korea and South Africa while the series are not co-integrated in USA (Jinke et al., 2008).

![Fig. 2.3: Relationship between coal consumption and GDP in key OECD and non-OECD countries (Jinke et al., 2008)](image)

Ozturk & Akaravci (2010) investigated the cause-and-effect relationship between energy and economic growth by using the variables of energy consumption per capita, electricity consumption per capita and real gross domestic product per capita. The time period studied by them was from 1980 to 2006 and they conducted their research in Albania, Bulgaria, Hungary and Romania as shown in Fig. 2.4 (Ozturk & Akaravci, 2010).

![Plot of Cumulative Sum of Recursive Residuals](image)

(a) Albania

![Plot of Cumulative Sum of Squares of Recursive Residuals](image)

(b) Bulgaria
Fig. 2.3: Stability parameter of GDP–EC1 equation (Ozturk & Akaravci, 2010)

Chang (2010) used multivariate cointegration test to examine the correlation between greenhouse gas production (carbon dioxide), energy consumption and economic growth in China. Results showed the Intensity of energy and energy-related carbon dioxide as shown in Fig. 2.5.

Fig. 2.5: Intensity of energy and energy-related carbon dioxide (Chang, 2010)

Narayan et al. (2010) tested the zero hypothesis about energy consumption in the countries of the Australian continent. They found that there was considerable evidence that imposed energy consumption shocks have a temporary effect on energy consumption in Australia (Narayan et al., 2010). They obtained time series plots for Australia, New South Wales as shown in Fig. 2.6 and Fig. 2.7 respectively.
According to the literature, it can be said that significant efforts have been made to predict energy consumption and demand. Box Jenkins models, regression, econometrics and NN are the most common techniques used in energy forecasting studies. Energy consumption forecasting and demand modelling is usually based on past consumption and the relationship between this consumption and other variables such as economic, demographic,
weather and energy price variables. The authors will now examine a number of research that have been conducted to model and predict energy consumption using various methods.

Dahl & McDonald (1998) developed a predictive model based on country-specific referentiality and an analysis of 25 countries around the globe.

Ediger & Tatlidil (2002) presented a method that uses the analysis of periodic patterns of the past curve for the purpose of preliminary forecasting of Turkey's energy demand.

Yumurtaci & Asmaz (2004) presented a method to calculate Turkey's future energy demand for the period 1980-2050 based on population and per capita energy consumption growth rate.

Mohamed & Bodger (2005) investigated the effect of economic and demographic variables (gross domestic product, average electricity price and population) on annual electricity consumption in New Zealand.

Using linear and non-linear statistical models, including the NN method, Pao (2006) investigated the effect of four economic factors: national income, population, gross domestic product, consumer price index on Taiwan's electricity consumption, and then developed an economic forecasting model.

Bianco (2009) investigated the effect of economic and demographic variables on the annual electricity consumption in Italy and developed a long-term consumption forecasting model. The studied economic and demographic variables were past electricity consumption, gross domestic product, gross domestic product per capita and population (Bianco, 2009).

Geem & Roper (2009) proposed an ANN model that had four independent variables such as GDP, population, import and export rate, which can effectively estimate South Korea's energy demand.

Ekonomou (2010) presented an ANN model including annual ambient temperature, installed electricity capacity, annual residential electricity consumption and GDP that predicts long-term energy consumption in Greece.

Soldo (2012) reviewed the articles in the field of forecasting the consumption and production of natural gas from 1949 to 2010. Articles were discussed in terms of application area, forecast horizon, input data and models. Brief list of studied researches has been shown in Table 2.1.

<table>
<thead>
<tr>
<th>Year</th>
<th>Researchers</th>
<th>Subject</th>
<th>Studied Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>Glasure &amp; Lee.</td>
<td>The case of South Korea and Singapore</td>
<td>Relation of energy consumption and GDP</td>
</tr>
<tr>
<td>1999</td>
<td>Ranjan &amp; Jain</td>
<td>electrical energy consumption pattern in India</td>
<td>Relation of consumption pattern of electrical energy by population and weather</td>
</tr>
<tr>
<td>2000</td>
<td>Yang</td>
<td>Study on relation of energy consumption and GDP in Taiwan for the period 1954–1997.</td>
<td>Relation of energy consumption, including coal, oil, natural gas and electricity by GDP</td>
</tr>
<tr>
<td>Year</td>
<td>Authors</td>
<td>Study Details</td>
<td>Relationship</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>2003</td>
<td>Soytas &amp; Sari</td>
<td>Study of relationship between energy consumption and income in G-7 countries excluding China due to lack of data</td>
<td>relation of energy consumption and income</td>
</tr>
<tr>
<td>2005</td>
<td>Narayan and Smyth</td>
<td>Study the relationship between electricity consumption, employment and real income in Australia</td>
<td>relation of electricity consumption, employment and real income</td>
</tr>
<tr>
<td>2006</td>
<td>Sözen &amp; Arcaklioğlu</td>
<td>Forecasting net energy consumption using artificial neural network</td>
<td>Relation of population, gross generation, installed capacity and years by energy consumption</td>
</tr>
<tr>
<td>2008</td>
<td>Jinkle et al.</td>
<td>Coal consumption and GDP relation in most OECD member and non-member countries</td>
<td>Relation of Coal consumption and GDP</td>
</tr>
<tr>
<td>2010</td>
<td>Ozturk &amp; Akaravci</td>
<td>Research in Albania, Bulgaria, Hungary and Romania from 1980-2006</td>
<td>Relationship between energy and economic growth</td>
</tr>
<tr>
<td>2010</td>
<td>Chang</td>
<td>Study on Energy consumption includes crude oil, coal, natural gas and electricity in China</td>
<td>Relation of carbon dioxide emissions, energy consumption and economic growth</td>
</tr>
<tr>
<td>2010</td>
<td>Narayan et al.</td>
<td>Energy consumption in Australia for the period of 1973–2007</td>
<td>Relation of shocks to energy consumption and energy consumption</td>
</tr>
</tbody>
</table>

Tamba et al. (2018) extended Soldo’s research up to 2015. They analyzed and synthesized published papers on models and application domain, input data, data source, data size, prediction horizon, results and model performance. Consumption statistics of natural gas in different regions has been presented by Enerdata (2020) as shown in Fig. 2.8.

![Fig. 2.8. Consumption statistics of natural gas in different regions (Enerdata, 2020)](image-url)
Development history of natural gas consumption forecasting has been studied by Liu et al. (2021) as shown in Fig. 2.9.

In 2006, Hinton developed a multi-layer NN and initially proposed the concept of deep learning, which represents machine learning research that is gradually entering the field of deep learning (Chen et al., 2018). The long-term short-term memory model (LSTM) is a special form of Recurrent Neural Network (RNN) and also the most popular deep learning model (Ghasemi et al., 2018) at present. This model can effectively solve the gradient vanishing problem of RNN, and the information of the previous moment in each cell is simultaneously recorded through the state variable, making the information relationship between different moments closer. Therefore, it is widely used to solve various time series forecasting problems (Graves, 2012; He, 2017).

By reviewing the studies, it can be seen that in most studies four independent variables of economic indicators and population have been used in order to model and predict energy consumption. As seen in the above studies, each of them have investigated and predicted energy using a specific method. These studies are divided into two groups: studies conducted with econometric methods and studies conducted with artificial intelligence (AI) and NN. In this research, an attempt is made to compare the results of the application of
NN for prediction in the field of energy with the regression model (RM) in order to compare the efficiency of this method with other methods. Also, at the end, the combined model of GA and NN is used to model and predict energy consumption, and its results were compared with NN and regression to determine the effectiveness of this combined model compared to other forecasting models.

3. Method of research
This research has several steps that are presented in the form of a flowchart in Fig. 3.1.

3.1. Step 1: Collecting data
Machine learning, a pivotal subset of artificial intelligence, is increasingly being recognized for its ability to develop algorithms that enable computers to learn from data and make informed predictions or decisions. This approach is fundamentally different from traditional programming, as it relies on the algorithm’s ability to identify patterns and make decisions with minimal human intervention. This concept is extensively explored in the field of pattern recognition and machine learning (Bishop & Nasrabadi, 2006). The application of machine learning in forecasting, especially in predicting heat usage in district heating networks, is gaining significant attention. Machine learning models, such as NN, RM, and more advanced techniques like the XGBoost algorithm, have shown considerable promise in analyzing complex datasets to predict future trends. This is particularly relevant in scenarios...
involving intricate variables like population growth, economic indicators, and building efficiency. For instance, Renuke et al. (2023) demonstrated the effectiveness of various supervised machine learning algorithms, including Linear Regression, Ridge, Gaussian Process Regressor, Random Forest Regressor, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks, in adaptive heat load consumption forecasting. The process of applying machine learning in forecasting involves several key steps: data collection, data preprocessing, model selection, model training, model evaluation, and prediction. This process is crucial in achieving accurate forecasting results, as evidenced by the work of Parrado-Duque et al. (2021), who conducted a comparative analysis of machine learning methods for short-term load forecasting systems. One of the significant advantages of using machine learning in forecasting is its ability to handle large, complex datasets, leading to more accurate and efficient predictions. This adaptability is essential in dynamic environments like weather and economic forecasting. For example, Bujalski et al. (2021) used the XGBoost algorithm for heat demand forecasting in a district heating network, highlighting the model's ability to incorporate various predictors for enhanced accuracy. In conclusion, the integration of machine learning in forecasting, particularly in the context of temperature and heat usage forecasting, represents a substantial advancement in predictive analytics. This approach not only offers a more nuanced understanding of future trends but is also crucial for effective planning and decision-making in various sectors.

3.2. Step 2: Pre-Processing
1. Initially, the outlier data were removed from the data set. Of course, to delete this data, you must have full confidence in the data. If the data is collected and only the statistical results are enough, it is better to remove the outlier data.
2. Subsequently, the data must be normalized. Therefore, the input and output data were normalized using Eq. 3.1.

\[ y = 0.8 \times \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} + 0.1 \]  

Eq. 3.1

3.3. Step 3: Training and testing
Selection of two sets of training and test data for the types of models used. Training data refers to the subset of data used to train a model. This data allows the model to learn the underlying patterns and relationships between the input variables and the target variable (Hastie et al., 2009). Essentially, the model uses this data to understand and make predictions. Test data, conversely, is a separate subset of data unseen by the model during the training phase. It's employed to evaluate the model's performance and accuracy in generating predictions based on patterns it learned from the
training data. For this study, distinct sets of training and test data were selected for the various models utilized to ensure their validity and robustness.

3.4 Step 4: Modelling and forecasting
Modelling and forecasting energy consumption using different methods.

1. By using the regression method step by step, the effective parameters for the models were identified and the assumptions of the regression method were also examined. Furthermore, the RM was fitted on the training data and after obtaining the desired equation, this equation was checked on the test data.

2. The second model under investigation is the neural network model (NNM). In this model, first the network is trained using the training data and then predicts the test data.

3. The third model used is the combined model of GA and ANN. In this model, the weights are first used by the optimal GA and then entered into the NNM so that the network is trained.

3.5 Step 5: Models Evaluation
In order to compare the performance of ANN-GA, artificial neural network and multivariate regression models(MRM), the mean error (ME), root mean square error (RMSE) and relative standard error (RSE) parameters were used, which can be calculated from the following equations. In addition to the aforementioned statistical criteria, the error square root percentage reduction index was also used to check the efficiency of different models (Relative Improvement, RI). In this model, RMSE_a is the square root of the error in the regression method and RMSE_b is the square root of the error in other methods. This statistic shows the amount of error reduction in different models compared to the regression method. In fact, it is a more suitable model that has more RI (Amini et al., 2005).

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(Z_0 - Z_p)^2}
\]

Eq. 3.2

\[
ME = \frac{1}{n} \sum(Z_0 - Z_p)
\]

Eq. 3.3

\[
RI = \left(\frac{RMSE_a - RMSE_B}{RMSE_a}\right) \times 100
\]

Eq. 3.4

\[
RSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(Z_0 - Z_p)^2}
\]

Eq. 3.5

In which:

Z_0 is Predicted values. Z_p is Observational values. Z_avg is the average of observed values and n is the number of data.
4. Method of Analysis

4.1. Artificial neural network

Artificial neural network (ANN) was chosen as the data analysis method. The ANN is a simulation method that is inspired by the study of the brain system and the NN of living organisms. The high-performance power of biological systems is due to the parallel nature of their neuronal programming. An ANN performs this structure by distributing the simulation in small and simple interconnected processing units called neurons. The main role of a biological neuron is to sum up its inputs until the sum of its inputs does not exceed the threshold, and then produce an output. Creating a good NN for a specific application is very important. Creating an optimal network includes choosing a suitable architecture, the number of layers, the number of units in each layer and the connections between units, choosing the functions to convert intermediate units to simple units, designing the training algorithm, choosing the initial weights in a special way, and the stopping rule (Gorr & Nagin, 1994). One of the most commonly used NN is the multilayer perceptron NN as shown in Fig. 4.1. A multilayer perceptron is a standard combination of inputs, linear and non-linear neural units, and outputs. The output of all processing units from each layer is transferred to all processing units of the next layer. The processing units of the input layer are all linear, but in the hidden layer of neurons with sigmoid tangent function. Hyperbolic or any other non-linear and continuously differentiable function can be used. Usually, to increase the training speed, the neurons of the linear output layer are selected.

![Perceptron structure with a hidden layer](image)

**Fig. 4.1:** Perceptron structure with a hidden layer

The main problem in these networks is to determine the number of hidden layers and the number of their neurons. The number of hidden nodes is

21
important because hidden nodes play a significant role in the nonlinear configuration properties of neural networks (Zhang, 2003). In determining the number of input nodes, using the trial and error method is the most useful, but in general, the number of neurons in the input layer represents the number of input variables.

4.2 Details of Neural Network

Artificial neural network, or neural network for short, is a computational tool inspired by the human brain. The human brain with a weight of about 1411 grams and a volume of over 1251 centimetres is one of the most amazing structures that we face in the creation system. This ultra-advanced system has the ability to use different systems within the organism, cause the organism to move, interpret and interpret the received information, and store all the required information so that it can be quickly accessed. As a result, it was possible to solve the problem. In addition, humans can solve problems by talking to others, interacting and providing appropriate solutions. Explaining how the brain works is one of the major issues facing science today. Although in recent years, due to the alignment of experimental and theoretical research, good progress has been made in the field of understanding how the brain works, but there is still a long way to go to fully understand how the brain works. The human brain consists of approximately 151 billion nerve cells (neurons), most of which are involved in information processing. Considering that each nerve cell receives up to 15,000 inputs from other cells, it is simply not possible to know how this complex structure works, and thus far, limited information has been obtained on the functioning of this huge structure. Usually, nerve cells communicate with their nearby cells and as a result, they become a cell complex. Although people's brains are different in size and shape like their faces, but the brain of all people has the same structure. About the structure of nerve cells, it can be said that nerve cells are similar to other body cells in many aspects. However, in some cases, they have special features that can achieve specific actions. Nerve cells do not have the ability to reproduce and repair after the transformation period, and there is no direct connection between them, but they are distinguished from each other by the charge space and their communication takes place through the chemical substances released in these spaces. The shape and size of nerve cells are variable, and this difference shows the adaptation of each nerve cell to its specific action. Nerve cells are able to change their behaviour as a result of experience and learning, and if they become dysfunctional, they create behavioural disorders. Unlike the nerve cells of the central nervous system, the nerve cells outside it, such as sensory cells (receptors) and motor cells, can be repaired. Each nerve cell consists of a membrane and a cell body, whose outer surface is covered with spiny dendritic processes. Dendritic spines are related to the synaptic terminals of other neurons. Each nerve cell has a long appendage called axon, the free side of which ends in terminal
branches (nerve terminal). Often, axons have a protein-fat coating, and the speed of the nerve current is directly related to the diameter of this coating. The end of the axon has many branches and several branches may be separated from it at the beginning. In general, input signals are received by the cell body and dendrites, and the cell body emits output signals after combining them (or in approximate language, averaging them), in addition, it is responsible for providing the life of the cell. The axon carries the output message from the cell body to the axon terminals and they transmit the information to other neurons. The messenger has two aspects: electrical and chemical. The message that is generated in the nerve cell and travels the length of the axon is an electrical impulse, but the transmission of the message from one cell to another is carried out by molecules of carrier substances, which pass through the synapse. The synapse is the distance between the sender of the message (axon terminal or sometimes dendrites) and the receiver of the message (dendrite or cell body and sometimes axon). A neuron usually receives messages from hundreds or thousands of other neurons and sends messages to hundreds or thousands of other neurons. A synapse is a basic structure and a functional unit between two neurons (the axon conduit of one neuron and the dendrite of another neuron). When the stimulus reaches the synapse terminal, a certain chemical interaction causes the transmitter of the neuron to be stimulated. Neuron impulses spread across the synaptic cleft, and depending on the type of synapse, the neuron receptor stimulates or inhibits the transmission of electrical signals. Synaptic connections can be modified by passing signals through it, so that synapses can learn from the activities they participate in. This historical dependence acts as memory in synaptic connections and provides the possibility of responding to human memory. A view of a human brain cell has been shown in Fig 4.2.

Fig. 4.2: View of a nerve cell (science photo library, 2023)
Work on AI began in the 1950s by pioneers in the fields of statistics, neuroscience, and psychology. In these kinds of methods, mankind seeks to conquer the universe and tries to use the best and most effective natural methods. One of the most important areas of AI is ANN, which seek to simulate the function of the small human brain in mastering the universe (Zahedi, 1993). The work on ANN or in general NN specifically started in 1943 by McCulloch and Pitts (McCulloch & Pitts, 1943). Since the purpose of AI is to develop paradigms or algorithms used by humans for use in machines, ANN, as one of the methods of AI, seek to imitate the functioning of the human brain. Following that, in recent years, we have seen a continuous movement from purely theoretical research to applied research, especially in information processing for problems that either do not have a solution or cannot be solved easily. Due to this fact, there is a growing interest in the theoretical development of model-free intelligent dynamic systems that are based on experimental data. NN (neural computations), fuzzy logic (approximate computations) and GA (genetic computations) are important and basic components of computational intelligence, each of which is modelled on the brain in some way. NNM synaptic communication and internal structure, fuzzy logic, approximate inference and genetic computation, mutational computation (Fausett, 1994). ANN are part of this group of dynamic systems that, by processing experimental data, transfer the knowledge or law hidden behind the data to the network structure. That's why these systems are called smart; because they learn general rules based on calculations on numerical data or examples. These systems based on computational intelligence try to model the neuro-synaptic structure of the human brain. Of course, it must be said that ANN are not comparable to the natural nervous system, and despite the exaggerations that are made about these networks, these networks do not try to preserve the complexity of the brain at all (Hollnagel, 1989). The first artificial neuron was presented by McCulloch and Pitts in 1943 (McCulloch & Pitts, W, 1943). Presented neuron of McCulloch and Pitts present a threshold unit as a computational model. Artificial neurons add the signals received from other neurons after multiplying by communication (synaptic) weights. The result of this addition passes through the activity function; the output of the activity function leads to the excitation or inhibition of the neuron. The artificial neuron is very similar to the biological neuron. It models internal wires and connections, or in other words, inputs and outputs, dendrites and axons, communication weights represent synapses, and the activity function estimates the function of the cell body. The similarities between biological neural networks (BNN) and ANN are:

1. Both ANN and BNN have a parallel structure.
2. The structural blocks of both networks have very simple computing devices.
3. The connections between neurons in both networks determine the performance of the network.

The differences between these two networks are:

1. The neurons of ANN are much simpler than the neurons of the BNN.
2. Artificial neurons that are made by electric circuits work much faster (about a million times) than biological neurons.
3. The structure of the brain is much more complicated than the structure of ANN. So that despite the higher speed of artificial neurons, brain performance is much faster than the performance of a normal computer. The cause of this phenomenon is the completely parallel structure of neurons. In such a way that all neurons usually fire and respond simultaneously

Complex conceptual decisions, such as face recognition, are made in the human brain in less than one hundred thousandth of a second. These decisions are made by a network of neurons whose operation speed is only a few thousandths of a second. This implies that the calculations cannot have more than about 100 serial steps. In other words, the brain executes a series of parallel programs that have about 100 steps for such a conceptual task. This issue is known as the hundred-step rule (Jain et al., 1996).

An ANN is a collection of interconnected neurons in different layers that send information to each other. The simplest form of network has only two layers: Input and output layer. The network acts like an input-output system and uses the value of the input neurons to calculate the value of the output neuron. Figure 4.3 shows a standard diagram of a neural network.

![Fig. 4.3: A standard diagram of a neural network.](image)

Each neuron is represented by a circle and the connection between neurons is represented by an arrow. The output \( y \) and the inputs \( X_0, X_1, X_2 \) are \( 1 \times n \) vectors where \( n \) is the number of observations. In this example, information moves exclusively from inputs to outputs, so the model in question is known
as a feed forward neural network (FFNN). The relationship between an input and an output is defined by a weight $a$, which indicates the relative importance of the said input in calculating the value of the output. In this way, the value of the output neuron of the observation input is obtained from the following relationship:

$$Net_t = a_0x_{0t} + a_1x_{1t} + a_2x_{2t} = \sum_{i=0}^{2} a_ix_{it} \quad \text{Eq. 4.1}$$

The output neuron then processes the obtained value using a transformation or activation function (engine) denoted by $f$. In the simplest form, the FFNN is a linear activation function, for example $f(x) = x$. The value obtained from the zero relationship of a linear activation function makes the final output of the network for observation $t$ as follows:

$$y_t = f(Net_t = a_0x_{0t} + a_1x_{1t} + a_2x_{2t}) = a_0x_{0t} + a_1x_{1t} + a_2x_{2t} \quad \text{Eq. 4.2}$$

Usually, one of the inputs has a value of one for all observations, and it is called a bias term. If we accept that $x_0$ is a skew term, then the output of the network is obtained from the following equation:

$$y_t = a_0 + a_1x_{1t} + a_2x_{2t} \quad \text{Eq. 4.3}$$

As can be seen, a FFNN with two layers and a linear activation function is similar to the multivariate linear regression model. Input neurons are independent variables or regressors and the output neuron is the estimate of the dependent variable. The different weights of the network are also similar to the parameters of the RM and the skew term is also the same width as the origin or the constant term in the RM. If we add the dependent variable intervals to the set of inputs, then we get a network similar to the linear autoregressive model. In general, the role of neurons in NN is to process information, and this is done in ANN by a mathematical processor, which is the activation function (AF). The AF can be linear or non-linear. An AF is selected by the designer based on the specific needs of the problem to be solved by the NN. For example, when the output values of the problem are only zero and one, it is no longer appropriate to use a linear activation function and one should use other functions that only result in zero and one values based on different input values (for example, the threshold function). Almost all NN use nonlinear AF in parts of the network. This allows the network to generate suitable nonlinear patterns from complex data sets. Ideally, the AF should be continuous, differentiable and uniform, because this problem facilitates the process of finding the appropriate coefficients of the optimization algorithm. The most common AF used in the NN literature is the logistic cumulative distribution function or the sigmoid function:

$$f(x) = \frac{1}{1+e^{-x}} \quad \text{Eq. 4.4}$$

The value of the logistic function is in the range of zero and one. NN researchers have tried to reproduce the activation state of real neurons in the
human brain by using bounded functions. When the function is close to one, the neuron is very active towards the received signals. When the function is close to zero, the neuron rarely reacts to the received signals.

If the variable we want to predict can also take negative values, then it is better to use the hyperbolic tangent activation function:

\[ f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \]  

Eq. 4.5

The hyperbolic tangent function has the same characteristics as the logistic function, but its value changes between -1 and 1. In the following, learning in NN and its types will be discussed.

Among the properties of the NN, the learning property of the network is of particular importance, which is discussed in this section. In the NN literature, instead of the term coefficient estimation, the term learning or training is used to find the value of network weights. As learning systems, NN have the ability to learn from past experience and environment and improve their behaviour while learning. Improvement in learning over time should be measured against a standard. The improvement measure models the same goal as the learning system. Here, the learning law is generally expressed in the form of differential equations by recursive relations. These recursive relations are called learning algorithms. In the learning process, NN adjust the synaptic weights to respond to the stimulus input, so that the output of the network converges to the desired output. When the actual output becomes the desired output, the training of the network is terminated, and the so-called knowledge network has learned. The purpose of the learning algorithm is to train the NN to perform a specific task. In other words, during training, after each repetition of the learning algorithm, NN become more informed about the environment, conditions and purpose of their work, and the type of learning is determined by the process according to which the parameters of the network are adjusted. Usually, the learning process starts with random selection of weights. The difference between the actual output (y) and the desired output (Z) is called delta. The goal here is to minimize the delta (or even better, make it zero). Delta reduction is done by making gradual changes in the weights. Here, to learn a neuron, the environment of the neuron is considered fixed. But when a NN changes and improves its behaviour collaboratively, not independently as we have seen for a single neuron, each vector neuron changes its corresponding weights according to its own learning rule and the information source environment. Each neuron in this state is no longer fixed but changes by changing the weights of other neurons because the information source environment of a neuron is not intrinsic but depends on the behaviour of other neurons in the network. Therefore, we can write the equations for the neurons of a network. In the following, the types of learning in NN are expressed more specifically.
1- Learning with a supervisor

In this teaching method, network inputs with corresponding outputs are already known. During training, input is applied to the network and the network produces an output in response to that stimulus input. This output is compared with the desired output, which is called the target output. Now, if the desired output differs from the actual output, the network generates an error signal, which is used for the amount of change that should be applied to the weights. This error minimization process requires a special circuit called the teacher or supervisor to perform the comparison between input and output. In supervised learning, the learning rule is given a set of data pairs, called learning data, where \( p_i \) is the input from the network and \( t_i \) is the desired output of the network for the input \( p_i \). After applying the input \( p_i \) to the NN, the parameter \( a_i \) is compared with \( t_i \) in the output of the network and then the learning error is calculated. Then that error is used to adjust the parameters of the network in such a way that if the same input \( p_i \) is applied to the network next time, the output of the network will be closer to \( t_i \). The degree of closeness is generally measured by the second power of the difference of vectors.

Fig. 4.4 illustrates well the learning with the supervisor. The teacher is a system that is aware of the environment (for example, it knows that for the input \( p_i \), the desired output is \( t_i \)). Note that the environment is unknown for the NN. At moment \( k \), the input vector \( p_i(k) \) with a certain probability distribution function, which is unknown to the NN, is selected and applied simultaneously to the NN and the teacher. The desired answer \( t_i(k) \) is also given to the NN by the teacher. In fact, the optimal response is the optimal response that the NN should reach for the given input.

Fig. 4.4: The learning with the supervisor
4.3 Optimization by Genetic Algorithm

Genetic algorithm (GA) was first proposed by Holland in 1975 and was developed by other researchers in the following years. GA is a part of the theory of evolutionary computation, which is currently growing rapidly as a part of AI. The main idea of this algorithm lies in Darwin's theory of evolution. From the practical point of view, the GA is one of the optimization methods, which is based on natural selection (the main factor of biological evolution) and some important concepts of genetic science. In this approach, the optimization of the objective function begins with an initial set of solutions, known as the initial population of chromosomes. Each chromosome represents a potential solution to the problem at hand. Through an iterative process, this initial population evolves into a new set of chromosomes, termed the new generation, which comprises refined solutions that are closer to the optimal answer for the given problem. By repeating this operation and generating a new population from the previous population at each stage and as a result of reaching successful generations, the population will grow towards an optimal response.

4.3.1 Implementation

In the following, the implementation details of the GA are described:

(a) Showing fields

The appropriate display of strings depends on the characteristics of the search space; but usually, they are shown as binary strings. In this article, the variables are coded in binary form with fixed string length. The strings used in the GA can be coded discretely or continuously. Due to the discrete nature of the variables used, in this research, each bit of the chromosomes of each generation represents one of the variables used; In this way, in each chromosome, zero bit means the absence of the corresponding variable and one means the presence of the corresponding variable in the final selected combination.

(b) Calculation of fitness

The fitness function (FF) is obtained by applying the appropriate transformation on the objective function that is to be optimized. This function evaluates each string with a numerical value that determines its quality. The higher the quality of the answer string, the higher the fitness value of the answer and the probability of participation for the generation of the next generation increases. In the problem of this research, the value of the FF is set equal to the inverse of the errors resulting from NN training for each field; This means that each string that represents a combination of variables is entered into the corresponding NN, the network is trained with the
corresponding data, and finally, the training error of the network is calculated. The inverse of this calculated error is equal to the FF of the GA.

(c) Population size

Goldberg in 1989 suggests sixty strings to calculate the best population size for binary codes of continuous variables of maximum length (Goldberg, 1989). In the following, GA operators will be examined.

1- Selection:

After the fitness of all the people of a generation is determined, according to natural principles, children born from more fit couples have more fitness. Just as in nature, people who have advantages over others achieve superior couples, the GA also simulates this process and gives fitter people more chances to reproduce. The easiest way to choose is to use the Roulette wheel as shown in Fig. 4.5.

![Roulette wheel](image)

**Fig. 4.5:** Roulette wheel (Goldberg, 1989)

In this method, a wheel with unequal segments is considered so that each person has a segment with a central angle. Now considering a random number in the range \([0,2\pi]\), a string is selected in which the random number is placed in its corresponding sector. So that the bigger sectors have more chances to be selected. Therefore, the first operator of the GA, i.e. selection, was investigated (Goldberg, 1989).

2-Crossover

Crossover is the most important operator of the GA and the key to its success. The selection operator does not have a tool to discover new areas of the search space, and if it is limited to copying the old structures without changing them, it is not possible to investigate new ones. Crossover is an operator that randomly exchanges information between threads. The simplest form of this
operator is a one-point Crossover. In the simple Crossover, first, a random number is chosen. Then, the corresponding bits are exchanged in two strings that must be combined together. In this way, two children are born as shown in Fig. 4.6. The Crossover operator can be two-point. In this case, two points are randomly selected and the Crossover action takes place between them.

<table>
<thead>
<tr>
<th>Chromosome 1</th>
<th>11011</th>
<th>00100110110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 2</td>
<td>11011</td>
<td>11000011110</td>
</tr>
<tr>
<td>Offspring 1</td>
<td>11011</td>
<td>11000011110</td>
</tr>
<tr>
<td>Offspring 2</td>
<td>11011</td>
<td>00100110110</td>
</tr>
</tbody>
</table>

Fig. 4.6: An example of single point crossover

Another method of crossover is uniform crossover. If all the chromosome points are selected as recombination points as shown in the Fig. 4.7, the recombination will be uniform.

Fig. 4.7: An example of Uniform Crossover

3- Mutation

The third important operator in the GA is called mutation. Although the selection and crossover search operators are effective in searching the design space, they sometimes cause us to lose the useful properties of strings. The mutation operator makes it possible to regain these positive characteristics that are not in the population. The behaviour of the mutation operator, in a simple way, is that for each individual, in the set (usually after the crossover operation), the probability of mutation occurrence, which is usually less than two percent, is checked. If a mutation is to be made, a point in the chromosome is chosen randomly, and its value changes from zero to one or vice versa. Fig. 4.8 shows an example of this operator.

Fig. 4.8: An example of Mutation operator
4- Convergence

The general optimization problem is unsolvable in its general state, so it cannot be expected to achieve the overall optimality of the function in a limited time. However, we are usually interested that, as a reassuring factor, our optimization algorithm converges to the overall optimality of the function with a probability of one. Rudolph (1994) has investigated the behaviour of simple GA in terms of convergence. His analysis shows that the simple GA does not converge to its overall optimum in infinite time (Rudolph, 1994). This result is not very worrying because, first, the algorithm is always executed in a limited time, and as a result, we usually have to settle for the general optimal approximation. Secondly, this result does not mean that the algorithm never reaches its overall optimum, but it does not mean that the algorithm reaches the overall optimum and leaves it. In fact, on average, this event occurs in a limited time. In infinite time, the algorithm reaches the global optimum an infinite number of times and leaves it. The issue that is more important than the convergence of the algorithm at infinity is the time that must be spent for the algorithm to reach the overall optimum for the first time. The results have shown that if the best individual of the population is stored in a memory separate from the population during the execution of the algorithm, the algorithm will converge to the overall optimum in infinite time. The best person can be found and stored in the population in one of the two stages before or after selection.

5. Combined model of genetic algorithm and artificial neural network

Neural Network (NN) and the Genetic Algorithm (GA) are foundational methodologies in this study. By merging the capabilities of both, the authors aim to achieve enhanced predictive accuracy for Sweden's heat consumption. The steps to develop the integrated model are as follows:

**Step 1:** The size of the population in each generation and the maximum number of generations are determined in the first step, and in this step a random initial population is created.

**Step 2:** In this step, the shape of the ANN is determined using the values of the genes in each created population.

**Step 3:** The designed network is trained using normalized input data. After training the network, calibration is also done in this step.

**Step 4:** After making the prediction, the mean square error (MSE) criterion is calculated using the designed network. By calculating this criterion, the objective function of the problem, which in this research is to minimize the MSE, is determined.
**Step 5:** In order to create the next generation, operators such as genetic and evolutionary operators such as gene combination and mutation, as well as the roulette cycle are used to select the next generation in the GA. At this stage, elitism is also used, with the help of which some of the best of the population of the current generation are transferred to the next generation.

**Step 6:** In this step, the newly created population replaces the previous population to create a new generation. In this step, 1 is added to the generation number and the above steps are repeated until the generation number reaches its maximum value. All the steps are shown in the following flow chart.
Fig. 5.1: flowchart of combined model of genetic algorithm and artificial neural network step by step.
6. Analysis and Results

6.1. Data Collection

In this thesis, the Sweden's heat use in residential sector data was used as the output variable of the forecasting models, and the annual data of the gross domestic product, total population of the country, heat consumption, district heating (DH) were used as the input variables of the forecasting models. The time period of these variables are from 1983 to 2021. These data of DH, oil consumption, electric heating, natural gas and biomass consumption are shown in Fig. 6.1. As can be seen in Fig. 6.1, the trend of heat consumption in Sweden is a strong downward trend from 1983 to 2021 in total. The following graph, derived from the Swedish Energy Agency and Statistics Sweden, showcases the variation in heat consumption in multi-family houses (MFB) in Sweden spanning the years 1983 to 2021. It's crucial to keep in mind that The two main building types in Sweden are single-family house(SFH) and MFB (Savvidou & Nykvist.,2020). During the past 30 years the Swedish residential heat system has shifted from oil as an important fuel driven by factors such as the oil crisis in the 1970–1980s (Nässén, & Holmberg, 2005) and swapped by two energy systems, DH and electricity through resistive heating and heat pumps (Savvidou & Nykvist.,2020). More than 75% of the energy used by families for heating is met by these two technologies. Specifically, DH which generates around half of the heat in the building stock (Savvidou & Nykvist.,2020).

![Variation of Heat Consumption in Sweden for MFB from 1983 to 2021. (Swedish Energy Agency,2022)](image)

**Fig. 6.1:** Variation of Heat Consumption in Sweden for MFB from 1983 to 2021. (Swedish Energy Agency,2022)

The core concept of DH is to "use local fuel or heat resources that would otherwise be wasted"(Werner, S., 2013). A DH system consists of a system of pipes that connects the buildings within a town, neighbourhood or city so
that they can receive heat from several dispersed heat-producing units or from central plants (Lund et al., 2017). The primary worldwide findings regarding DH include low building utilization of the technology, uneven adoption rates, a moderate dedication to the core concept of the technology, and little recognition of potential reductions in carbon dioxide emissions (Werner, S., 2017). Waste-to-energy resources, combined heat and power plants, and industrial operations are examples of traditional sources of surplus heat (Werner, S., 2017). DH systems supply heat to around 50% of the nation's building stock in nations with solid economies. Due to a lack of knowledge or competition in DH, relatively few systems exist in countries with weak economies (Werner, S., 2017). The first heat distribution system employed steam, but in the second generation, high-temperature water that was above 100 °C replaced the steam (Werner, S., 2017). In the third and fourth generations, the water's temperature dropped to around 80 and 60 °C (Pellegrini & Bianchini, 2018). Fifth-generation district heating and cooling is a new generation that provides simultaneous heating and cooling to connected customers who are producers or consumers of thermal energy (Revesz et al., 2020). This development is a result of growing cooling needs brought on by climate change and urbanization. It should be noted that third-generation DH systems, which distinguish between the utility provider and the client, make up a vast majority of systems in use in Sweden today (Lygnerud, K. 2019). Since steam is regarded as an inefficient heat carrier in terms of heat losses and maintenance costs, the majority of first-generation systems have either been closed down or converted to water systems (Werner, S., 2017). Approximately 55% of all buildings' heat demand and up to 92% of multi-family residential dwellings' heat demand are met by DH (Åberg, M., 2014). The fuels and technology utilized to generate heat in Swedish DH systems are very diverse, and the majority of the heat is produced locally through the harvesting of waste heat resources (Romanchenko et al., 2017). It should be mentioned that heat is the primary product of DH, with electricity serving as a byproduct and this research's primary focus is on heat. According to Abugabbara et al., (2023) In 2030, DH in Sweden will be completely decarbonized, and by 2050, the country aims to have zero greenhouse gas emissions. The EU Strategy on Heating and Cooling views district heating and cooling systems as practical ways to help the construction industry become decarbonized (An, E. U., 2016). Deep decarbonization has a main goal which is to reduce greenhouse gas emissions by at least 80 percent by 2050. DH systems contribute to this aim and reduce the impact of climate change globally. While there are several ways to achieve deep decarbonization, practical routes are restricted to a limited number of technologies due to financial implications, technological limitations, and other variables (Holmes et al., 2021).
The changes of Sweden's population from 1983 to 2021 are shown in Figure 6.2. As can be seen in Fig. 6.2, the trend of population growth in Sweden is a strong upward trend from 1983 to 2021.

![Population Chart](image)

**Fig. 6.2:** The changes of Sweden's population from 1983 to 2021 (Macrotrends, 2023a)

The changes of Sweden's GDP from 1960 to 2022 are shown in Fig 6.3. As can be seen in this figure, the trend of GDP changes in Sweden is an upward trend and experienced recession in the years 1981 to 1983 and 1993 and the period of 1997 to 2002. It is obvious that Sweden, like other countries, was affected by the great recession in 2008.

![GDP Chart](image)

**Fig. 6.3:** The changes of Sweden's GDP from 1960 to 2021 (Sweden GDP 1960 to 2021) (Macrotrends, 2023b)
The changes of Sweden's DH supply from 1970 to 2022 are shown in Fig. 6.4. As can be seen in this figure, the trend of DH consumption changes in Sweden from 1970 to 2010 was an upward trend. It is noteworthy that since 2010, the trend of DH consumption changes has changed to a downward trend.

![Fig. 6.4: The changes of Sweden's District heating from 1970 to 2022. (Swedish Energy Agency, 2022)](image-url)
6.2. Results

Variation of district heating by year has been plotted as shown in Fig. 6.5.

In our work, the graphical representation as depicted in Figure 6.5 serves to illustrate the annual variations in DH usage, accompanied by a forecast model that extends to 2050. The observed fluctuations within the forecasted data can be primarily attributed to several interrelated factors, which are comprehensively delineated in our study. Examining the short-term trends leading up to the year 2025, we notice a temporary downward trajectory in DH usage. This could be interpreted as the result of anticipated advancements in energy efficiency, the cumulative impact of regulatory measures aimed at reducing overall energy consumption, or potentially, the prediction of milder weather patterns which would naturally lead to a reduced demand for heating. In contrast, the trend beyond 2025 shifts to an upward climb, which may signal an expected increase in heating demand. This anticipated rise could be driven by various determinants, including demographic growth, likely changes in climatic conditions that may result in cooler temperatures, and
other socio-economic factors that typically contribute to increased energy usage. The prowess of the neural network model (NNM) employed in projecting these trends is substantiated by its high accuracy rates. This is evidenced by the mean square error (MSE) values, which indicate a strong congruence between the model's predictions and the actual data. During the training phase, the model demonstrates a near-perfect correlation with actual targets, reflected by a 99% accuracy rate. While there is a nominal dip in the validation phase to 94% accuracy, and a further decrease to 87% accuracy in the testing phase, the model still maintains an overall average accuracy of 96%. This robust performance endorses the reliability of the model in its predictive capacity. Our research emphasizes the complex interplay of various parameters—such as heat consumption, temperature fluctuations, population growth, building insulation efficiency (U-value), and economic activity as measured by GDP—which are all pivotal in forecasting heating requirements. The model adeptly incorporates how shifts in one parameter can impact another, for instance, how variations in average temperatures can alter energy needs for heating. The innovative approach presented in our thesis involves a synthesis of multiple regression analyses with the neural network's optimization of variable parameters, further enhanced by GA. This sophisticated approach underpins the precision of our predictions regarding future heat consumption. In summary, the forecasted fluctuations are indicative of the dynamic interactions among several socio-economic and environmental factors that critically influence DH consumption. Our analysis presents an integrative perspective on these variables, offering a forward-looking approach to anticipating future energy demands. Performance of presented NN has been presented in Fig. 6.6. Mean Square Error (MSE) of performance of presented network has been achieved as 0.3530. Training State of the network and its histogram have been shown in Fig. 6.7 and Fig. 6.8 respectively.
Figure 6.6 presents the NN training interface of MATLAB's 'nntool'. The depicted NN consists of three layers:

1. An input layer with 7 neurons, representing the input data features.
2. A hidden layer with 18 neurons, serving as the computational layer where most of the weighted sums and activations take place.
3. An output layer with a single neuron, denoting the singular output the network produces.

The network utilizes the weights 'W' and biases 'b' to adjust and fine-tune its calculations during the training phase. For the training process:

1. The chosen algorithm is 'Levenberg-Marquardt' (often denoted as 'trainlm'), which is a popular choice for training NN due to its convergence properties.
2. Data division is noted to be 'Random', meaning data for training, validation, and testing is likely to be randomly divided, although exact percentages are not mentioned.
3. The primary metric for evaluating the network's performance during training is the MSE. Lower MSE values would typically signify better performance, and from the figure, we see a performance of 20.5 at
epoch 0, which indicates the initial error before any training iterations have taken place.

4. The 'Epoch' represents iterations or forward and backward passes of all the training examples. The progress bar showcases the network's journey from the 0th to the 1000th epoch.

5. Other notable parameters include the 'Gradient', which at 70.3, is a measure of how much the error will change as the network changes its weights and biases. A smaller gradient often suggests that the network is closer to finding an optimal solution. 'Mu' is an adaptive value used in the Levenberg-Marquardt algorithm to balance between the speed of convergence and the stability of the process.

The figure 6.6 serves as a visual representation of the network's architecture and an initial display of its training metrics. Figure 6.7 showcases the training state of the NN over 5 epochs, divided into three primary subplots:

1. Gradient: The top subplot displays the gradient's trend across the epochs. The gradient measures how much the error would change as the network adjusts its weights and biases. At the 5th epoch, the gradient is at a value of approximately $2.3388 \times 10^{-14}$, which is a very small value. This suggests that the network might be approaching an optimum since the gradient, which indicates the steepness of the error surface, is nearing zero.

2. Mu: The middle subplot represents the trend of the 'Mu' parameter across the epochs. 'Mu' is an adaptive value used in the Levenberg-Marquardt algorithm to balance between the convergence speed and the stability of the training process. At epoch 5, 'Mu' is at a value of $1 \times 10^{-8}$, showcasing the adjustment factor used in the optimization algorithm at that specific iteration.

3. Validation Checks: The bottom subplot indicates the number of validation checks performed during training. These checks are crucial as they help in preventing overfitting by monitoring the network's performance on a validation set. At epoch 5, there have been 3 validation checks, represented by the diamond-shaped markers. The increasing trend in validation checks suggests that the network might be closely monitored for its generalization capability.

The x-axis represents the 5 epochs or iterations, which are distinct forward and backward passes of the entire training dataset. Overall, Figure 6.7 provides insights into the training dynamics of the NN, capturing the essence of its optimization journey over the initial 5 epochs.
Figure 6.8 represents the Error Histogram of the NN performance, segmented across 20 bins. In the provided histogram, the term 'bins' delineates the discrete intervals that segment the continuum of error values into 20 distinct categories. Each bin encompasses a specific range of error magnitudes—quantified as the deviation between the NN predicted outputs and the target values. The histogram's bars reflect the frequency of instances within these error intervals, thereby offering a visual representation of the error distribution across the training, validation, and test datasets. Such a histogram is instrumental in evaluating the NN predictive accuracy, with the ideal distribution being one where the majority of instances are aggregated near the zero-error bin, indicative of minimal prediction error. The histogram visually depicts the distribution of errors (computed as the difference between targets and outputs) for the training, validation, and test datasets. Here's a breakdown of the key elements in the figure:

1. Training (Blue Bars): The blue bars in the histogram represent the error distribution for the training dataset. Notably, the majority of errors in the training data are concentrated around the -0.0093 and -0.0024 regions, indicating frequent minor discrepancies between the predicted outputs and actual targets during training.

2. Validation (Red Bars): The red bars provide insights into the error distribution for the validation dataset. These errors appear less frequent when compared to the training errors but are scattered across
various bins, highlighting the variances in the model's predictions when exposed to unseen validation data.

3. Test (Green Bars): The green bars show the error distribution for the test dataset. These errors are more spread out, with a significant cluster around the 0.0585 and 0.1612 regions. This may suggest that on specific test instances, the network's predictions deviated more substantially from the expected targets.

4. Zero Error (Black Line): The black vertical line signifies the 'Zero Error' mark. Any bars to the left of this line indicate under-predictions, whereas bars to the right represent over-predictions. This gives a clear visual cue about the overall bias in the model's predictions across different datasets.

The y-axis denotes the number of instances that fall within a particular error range, while the x-axis represents the error values, segmented into 20 distinct bins. Overall, Figure 6.8 offers a comprehensive view of the model's prediction accuracy and areas of improvement across training, validation, and testing phases. It emphasizes the model's tendencies and showcases regions where the errors are most and least frequent.

![Error Histogram with 20 Bins](image)

**Fig. 6.8**: Error Histogram

Figure 6.9 depicts the regression analysis of the presented NNM across various phases: Training, Validation, Test, and an overall view. The objective
of such plots is to evaluate how well the model's outputs (predictions) correlate with the actual targets. Here's a detailed breakdown:

1. Training Regression (Top-left quadrant with blue line):
   - R-value: 0.99092, indicates a high degree of correlation between the model's predictions and the actual targets during the training phase.
   - Data Points: Represented as circles, they showcase the predicted output values against the actual target values.
   - Fit Line (Solid blue): This line showcases the model's predictive fit. The closer this line is to the dotted line (Y=T), the better the model's predictions.

2. Validation Regression (Top-right quadrant with green line):
   - R-value: 0.93984, although slightly lower than the training R-value, it still represents a strong correlation in the validation phase.
   - Fit Line (Dashed green): Represents the model's predictions for the validation set.

3. Test Regression (Bottom-left quadrant with red line):
   - R-value: 0.87846, suggests a good correlation between targets and outputs in the test phase.
   - Fit Line (Solid red): Illustrates the model's predictive capability for the test dataset.

4. Overall Regression (Bottom-right quadrant with black line):
   - R-value: 0.96048, representing the overall performance of the model across all datasets.
   - Fit Line (Dashed black): Demonstrates the cumulative predictive capability of the model.

In each quadrant, the dotted line (Y=T) represents an ideal scenario where predictions perfectly match the targets. The closer the fit line is to this dotted line, the better the model's performance. In summary, Figure 6.9 provides a comprehensive understanding of the model's predictive performance across different datasets. It highlights the areas where the model excels and offers insights into potential regions of improvement. As can be seen in Figure, the accuracy of the NN was 99% in the training phase, 94% in the validation phase, 87% in the testing phase and 96% as an overall average, which shows a very good accuracy.
Regression of presented model also has been shown in Fig. 6.9. As can be seen in Figure 6.9, the accuracy of the NN was 99% in the training phase, 94% in the validation phase, 87% in the testing phase and 96% as an overall average, which shows a very good accuracy.

7. Innovation of the Thesis
With time changes, parameters of heat consumption, temperature, population, U-value and GDP of Sweden changes. The parameters - heat consumption, temperature, population, U-value, and GDP - were chosen to holistically encompass Sweden's energy landscape. Heat consumption measures energy needs; temperature indicates environmental impact; population signifies residential demand; U-value denotes building efficiency; and GDP mirrors
economic growth and consumption behavior. These factors collectively represent technological, structural, and behavioral aspects crucial for forecasting energy demands (Aziz et al., 2022). The important note is that these parameters affect each other. That is, the change of each parameter leads to the change of other parameters. For example, annual temperature changes in Sweden lead to changes in the fuel consumption of residents. Considering that in the current project, the goal of the project is to predict the country's heat consumption until 2050, and because until 2050, the parameters of the country's population, the country's temperature, the value of U and Sweden's GDP will change, so it is necessary to achieve a relationship that will help heat consumption to be predicted with the help of the mentioned parameters. The 7.1 formula is obtained based on multiple regressions and by combining the optimization of the variable parameters of the NN with the GA and establishes the relationship between the parameters of the country's population, the country's temperature, the value of U and the GDP of Sweden with fuel consumption.

\[
heat\ usage = \left( \frac{10^{12}}{population^2} \right) + (U\ value^2) + \left( variable\ *\ (temperature^2) \right) + \left( \frac{GDP}{10^4}\ *\ GDP \right)
\]

Eq. 7.1

In Which:

Relation between population, U-value, temperature and GDP of Sweden have been considered. The parameter “Variable” also has the relation by variation of temperature as shown in Fig. 7.1. This figure illustrates the relationship between temperature and the "Variable" parameter, which represents a composite of key factors like heat consumption, population growth, building efficiency (U-value), and economic indicators (GDP). The graph provides essential insights into how these combined factors, encapsulated in the "Variable," are influenced by temperature changes, thus aiding in comprehensively forecasting Sweden's future energy needs and understanding the multifaceted interdependencies within its energy landscape. Mentioning Fig. 7.1 is crucial because it visually elucidates the intricate interplay between temperature and the "Variable" parameter, which aggregates pivotal determinants such as heat consumption, population growth, building efficiency, and GDP. This depiction aids in understanding and anticipating the broader implications of temperature variations on Sweden's energy dynamics, underscoring the interconnected nature of these factors and their collective impact on energy forecasting. The critical role of sophisticated database management in the accurate prediction of residential heating demands cannot be overstated. This task involves the complex integration of heterogeneous data sources, encompassing extensive meteorological data, historical energy consumption metrics, and detailed characteristics of residential structures (Golmohamadi, H, 2022). The design of a database system that can handle such diverse data types is nontrivial,
especially when considering the dynamic nature of time series data—a common feature of energy consumption data—which may not be seamlessly handled by traditional relational database systems (Feng, H. 2012). To ensure that forecasting models remain robust over time, they must be designed with adaptability and scalability at their core. Such systems should be capable of integrating evolving data without significant interruptions or the need for comprehensive system overhauls (Noh & Rajagopal, 2013). The veracity of the data housed within these systems is of utmost importance; even minor inaccuracies can lead to substantial deviations in forecast results, with potential ramifications for energy management and policymaking (Luo & Oyedele 2021). Within this framework, the selection of the appropriate database management approach is pivotal. The Entity-Relationship (ER) model is renowned for its efficacy in mapping out conceptual designs, delineating entities, and the relationships between them. This contrasts with the relational model, which is lauded for its logical structure and efficient data retrieval capabilities. In choosing to represent temperature as a singular concept within the ER model, and deliberately excluding additional variables such as humidity and wind speed, there is an inherent trade-off. This reduction simplifies the conceptual design and can enhance the clarity and focus of the model. However, it may also lead to a loss of detail that could be pivotal for the granularity and accuracy of predictions. In the context of database design, the notion of "Variable" can be interpreted as a derived attribute, which is not stored in its raw form. Instead, it is computed from other base attributes within the database when needed. The rationale behind using derived attributes, such as the "Variable" in this context, is strategic. It minimizes the storage requirements and ensures that the most current data is used in computations, which is particularly important when the derived data is subject to frequent changes or requires complex calculations. This approach, however, introduces additional computational overhead during data retrieval, as the derived attributes must be calculated in real-time. Therefore, a balance must be struck between the efficiency of data storage and the computational demands of data retrieval. In conclusion, the choice between utilizing the ER and relational models, as well as the decision to implement stored versus derived attributes, are deliberate and strategic, directly affecting the efficiency, scalability, and accuracy of the forecasting system. The decision to represent temperature like ER model since it doesn’t have full details like humidity percentage and rain velocity and define "Variable" as a derived attribute in this context is a calculated approach that aims to streamline the database design, ensuring the forecasting system remains both responsive and accurate. The "Variable" equation was used for temperature forecasting because it combines multiple regression analysis and optimization techniques from NN and GA. This approach enables the establishment of a relationship between critical factors like population, temperature, U-value (which denotes building efficiency), and GDP with fuel consumption. Specifically, the
The equation is designed to factor in the variation of temperature and its impact on heat consumption, population growth, building efficiency, and economic indicators. This comprehensive approach allows for a more accurate and detailed prediction of heat usage, as it considers the interconnectedness of these variables and their collective influence on Sweden's heating demands. The inclusion of temperature in the "Variable" parameter is particularly crucial because temperature fluctuations significantly impact heat consumption patterns. By integrating temperature variations into the model, the forecasting becomes more dynamic and responsive to environmental changes, which is essential for accurate long-term predictions in the context of Sweden's energy needs.

For verification of predicted district heating and the real data, reader can see the Table 7.1.

**Table 7.1:** Comparison of prediction of DH by presented formula.

<table>
<thead>
<tr>
<th>Year</th>
<th>District heating</th>
<th>Population</th>
<th>U-value</th>
<th>Temperature</th>
<th>GDP</th>
<th>Variable</th>
<th>Prediction of Heat Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>15.6</td>
<td>8329033</td>
<td>0.27</td>
<td>4.61</td>
<td>1.0501E+11</td>
<td>0.63</td>
<td>15.6</td>
</tr>
<tr>
<td>1984</td>
<td>16.4</td>
<td>8336605</td>
<td>0.26</td>
<td>4.55</td>
<td>1.0920E+11</td>
<td>0.66</td>
<td>15.9</td>
</tr>
<tr>
<td>1985</td>
<td>19.5</td>
<td>8350386</td>
<td>0.25</td>
<td>2.03</td>
<td>1.1412E+11</td>
<td>4.15</td>
<td>19.5</td>
</tr>
<tr>
<td>1986</td>
<td>19.7</td>
<td>8369829</td>
<td>0.24</td>
<td>3.43</td>
<td>1.5049E+11</td>
<td>1.39</td>
<td>19.7</td>
</tr>
<tr>
<td>1987</td>
<td>21.7</td>
<td>8397804</td>
<td>0.23</td>
<td>2.55</td>
<td>1.8301E+11</td>
<td>2.66</td>
<td>21.7</td>
</tr>
<tr>
<td>1988</td>
<td>18.9</td>
<td>8436489</td>
<td>0.22</td>
<td>4.46</td>
<td>2.0698E+11</td>
<td>0.68</td>
<td>18.9</td>
</tr>
</tbody>
</table>

![Graph](image-url)  
**Fig. 7.1:** Variation of temperature and Variable parameters

\[ y = 0.2292x^2 - 2.6326x + 7.6709 \]

\[ R^2 = 0.7377 \]
As can be seen in the column of “Prediction of District heating” in Table 7.1, by presented formula we can compare the results by the column of “District heating” in this table that shows an exact match of results from 1983 to 2006 by using the values in column of “Variable”. So, by the presented formula, exact prediction of DH from 1983 to 2006 is possible.

Based on the data in table 7.2, after 2007 to 2021, because of the fast growth of GDP of Sweden, the Variable parameter should be considered as Zero and GDP is the main effective parameters of the presented model. Table 7.2 shows the variations from 2007 to 2021.

<table>
<thead>
<tr>
<th>Year</th>
<th>District heating</th>
<th>Population</th>
<th>U value</th>
<th>Temperature</th>
<th>GDP</th>
<th>Variable</th>
<th>Prediction of Heat Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>22.8</td>
<td>9148092</td>
<td>0.2</td>
<td>5.4</td>
<td>4.91253E+11</td>
<td>0</td>
<td>25.2</td>
</tr>
<tr>
<td>2008</td>
<td>22.3</td>
<td>9219637</td>
<td>0.2</td>
<td>5.61</td>
<td>5.17706E+11</td>
<td>0</td>
<td>27.8</td>
</tr>
<tr>
<td>2009</td>
<td>21.9</td>
<td>9298515</td>
<td>0.2</td>
<td>4.79</td>
<td>4.36537E+11</td>
<td>0</td>
<td>20.1</td>
</tr>
<tr>
<td>2010</td>
<td>24.9</td>
<td>9378126</td>
<td>0.2</td>
<td>2.93</td>
<td>4.98513E+11</td>
<td>0</td>
<td>25.6</td>
</tr>
<tr>
<td>2011</td>
<td>21.1</td>
<td>9449213</td>
<td>0.2</td>
<td>5.69</td>
<td>5.74094E+11</td>
<td>0</td>
<td>34</td>
</tr>
<tr>
<td>2012</td>
<td>23.3</td>
<td>9519374</td>
<td>0.2</td>
<td>4.43</td>
<td>5.52484E+11</td>
<td>0</td>
<td>31.6</td>
</tr>
<tr>
<td>2013</td>
<td>23.0</td>
<td>9600379</td>
<td>0.2</td>
<td>5.04</td>
<td>5.86842E+11</td>
<td>0</td>
<td>35.5</td>
</tr>
<tr>
<td>2014</td>
<td>22.0</td>
<td>9696110</td>
<td>0.2</td>
<td>6.21</td>
<td>5.81964E+11</td>
<td>0</td>
<td>34.9</td>
</tr>
</tbody>
</table>
8. Conclusion

In the initial chapter of this thesis, a detailed examination of the general aspects of the research was presented. The focal problem of the research was elucidated, emphasizing its significance and justifying the need for its investigation. Subsequent sections introduced the research questions, followed by the delineation of both the general and specific objectives of the study. A concise overview of the research methodology was then presented to provide clarity on the approach adopted. In the second chapter, an exploration into previous research was undertaken to identify potential gaps in the existing literature. This scrutiny ultimately paved the way for the central theme of this dissertation: forecasting the heating consumption within Sweden's domestic sector. The third chapter involved a comprehensive evaluation of various methodologies. It was deduced that NN offer a viable solution for forecasting the heating consumption within the Swedish domestic milieu. This prompted a deep dive into the intricacies of the NN method, with a particular emphasis on the application of ANN. To enhance the accuracy of the predictive model, integration with numerical optimization methods was explored. The GA emerged as a suitable companion to the ANN, warranting its inclusion in the analytical framework. Chapter four showcased the empirical findings derived from the research, presented through an array of graphs and tables. These visuals accentuated the shifts in parameters influencing heating consumption in Swedish homes. The primary takeaways from these results are as follows:

1. Heat consumption in Sweden has observed a notable decline from 1983 to 2021.
2. There has been a consistent surge in Sweden's population from 1900 to 2022.
3. Fluctuations in Sweden's GDP, including recessions in the early '80s, mid-'90s, and the significant 2008 downturn, underscore its dynamic nature.
4. Sweden's DH consumption recorded an increase from 1970 to 2010, after which it witnessed a decline.

<table>
<thead>
<tr>
<th>Year</th>
<th>Heat Consumption</th>
<th>Population</th>
<th>GDP</th>
<th>DH Consumption</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>22.1</td>
<td>9799186</td>
<td>0.2</td>
<td>5.91</td>
<td>0.0</td>
</tr>
<tr>
<td>2016</td>
<td>24.0</td>
<td>9923085</td>
<td>0.2</td>
<td>5.39</td>
<td>0.0</td>
</tr>
<tr>
<td>2017</td>
<td>24.0</td>
<td>10057698</td>
<td>0.2</td>
<td>5.15</td>
<td>0.0</td>
</tr>
<tr>
<td>2018</td>
<td>24.2</td>
<td>10175214</td>
<td>0.2</td>
<td>5.63</td>
<td>0.0</td>
</tr>
<tr>
<td>2019</td>
<td>23.6</td>
<td>10278887</td>
<td>0.2</td>
<td>5.52</td>
<td>0.0</td>
</tr>
<tr>
<td>2020</td>
<td>22.7</td>
<td>10353442</td>
<td>0.2</td>
<td>6.6</td>
<td>0.0</td>
</tr>
<tr>
<td>2021</td>
<td>24.6</td>
<td>10415811</td>
<td>0.2</td>
<td>5</td>
<td>0.0</td>
</tr>
</tbody>
</table>
5. Between 1983 and 2006, the "Variable" value was significantly influenced by other parameters, with temperature playing a pivotal role. The relationship between the two was inversely proportional.

6. The rapid GDP growth in Sweden from 2007 to 2021 altered this dynamic, diminishing the significance of the "Variable" value. The GDP emerged as the paramount factor in the model during this period.

7. A novel formula encapsulating the correlation between Sweden's DH, population, U-value, temperature, and GDP has been introduced in this study.

While the analysis captured various economic trends, it did not explicitly incorporate the time value of money. This aspect is vital for ensuring fair and precise comparisons across different timelines, and its inclusion can further refine the insights derived. The current scope of the thesis did not extend to forecasting future trends using the developed model. While predictive analytics were central to the study, future projections were not a part of the current research iteration. In conclusion, this thesis sheds light on the intricate dynamics of heating consumption in Sweden's domestic sector, offering valuable insights and introducing a novel predictive formula. Future endeavours can benefit from extending the model's predictive capabilities and incorporating elements like the time value of money for a more holistic analysis.


In the realm of national and global energy modelling, the application of machine learning (ML) has become increasingly crucial. Its capacity to handle and analyze complex datasets enhances the accuracy and efficiency of forecasts. This thesis delves into the modelling and forecasting of energy consumption in Sweden, with a focus on predicting heat usage in DH networks using ML techniques, a subject that has garnered significant attention recently. The study primarily focuses on variables such as population, GDP, U-value, and temperature to predict heat consumption up to the year 2050, particularly for MFH. While earlier research concentrated on electricity or general energy consumption, this study aims to create both linear and non-linear models to forecast heat consumption using a combination of GA and NN. This approach aligns with recent advancements in non-linear ML techniques for energy consumption forecasting in smart grids, as demonstrated by Bharathi & Rekha (2023). Energy consumption modelling is crucial for effective energy planning and policymaking. It involves understanding past and present consumption patterns and estimating future demand, considering various economic, demographic, and weather-
related variables. The thesis leverages advanced ML techniques, including ANN and GA, to forecast future heat consumption in Sweden, similar to the approach used by Renuke et al. (2023) in their study on ML-assisted adaptive heat load consumption forecasting in DH networks. The study underscores the importance of energy consumption modelling and forecasting for the development and progress of countries. Accurate energy consumption models are pivotal in avoiding potential energy outages or wasteful capacity building due to overestimation. The thesis explores different models to determine the most effective method for modelling energy consumption based on various indicators, akin to the methods used in electricity consumption forecasting in office buildings by Jozi et al. (2019). In conclusion, the integration of ML in forecasting, particularly in the context of temperature and heat usage forecasting, represents a substantial advancement in predictive analytics. This approach not only provides a more nuanced understanding of future trends but is also crucial for effective planning and decision-making in various sectors, especially in light of the challenges posed by energy demands and the transition to renewable energy sources. Incorporating this study into national energy models, such as the TIMES model, the LEAP system, or the World Energy Model used by the IEA, could significantly enhance the accuracy and reliability of energy demand forecasts. These models would benefit from the advanced predictive capabilities demonstrated in this thesis, as evidenced by the novel integrated ML techniques proposed by Eseye & Lehtonen (2020) for short-term forecasting of the heat demand of buildings.
References


63) Science photo library (2023). Figure 3-3: View of a nerve cell. Available at: https://www.sciencephoto.com


Appendix
1- MATLAB code for forecasting heat use by Neural Network

clc
clear
load District_heating

% Year
data_Year = (unnamed);

input_train=[data_Year(20:38);data_Year(19:37);data_Year(18:36);data_Year(17:35);data_Year(16:34);data_Year(15:33);data_Year(14:32)];
target_train=[data_Year(21:39)];

net=fitnet(18);
net=train(net,input_train,target_train);
y=net(input_train)
MSE_perf=mse(y-target_train)

2- MATLAB code for forecasting heat use by Genetic Algorithm and Neural Network

%%% Start of Program

clc
clear
close all

%%% Data Loading
Data = xlsread('Data.xlsx');
X = Data(:,1:end-1);
Y = Data(:,end);
DataNum = size(X,1);
InputNum = size(X,2);
OutputNum = size(Y,2);

%%% Normalization
MinX = min(X);
MaxX = max(X);
MinY = min(Y);
MaxY = max(Y);
XN = X;
YN = Y;
for ii = 1:InputNum
    XN(:,ii) = Normalize_Fcn(X(:,ii),MinX(ii),MaxX(ii));
end
for ii = 1:OutputNum
    YN(:,ii) = Normalize_Fcn(Y(:,ii),MinY(ii),MaxY(ii));
end

%% Test and Train Data
TrPercent = 80;
TrNum = round(DataNum * TrPercent / 100);
TsNum = DataNum - TrNum;
R = randperm(DataNum);
trIndex = R(1:TrNum);
tsIndex = R(1+TrNum:end);
Xtr = XN(trIndex,:);
Ytr = YN(trIndex,:);
Xts = XN(tsIndex,:);
Yts = YN(tsIndex,:);

%% Network Structure
pr = [-1 1];
PR = repmat(pr,InputNum,1);
Network = newff(PR,[5 OutputNum],{'tansig' 'tansig'});

%% Training
Network = TrainUsing_GA_Fcn(Network,Xtr,Ytr);

%% Assesment
YtrNet = sim(Network,Xtr)';
YtsNet = sim(Network,Xts)';
MSEtr = mse(YtrNet - Ytr)
MSEts = mse(YtsNet - Yts)

%% Display
figure(1)
plot(Ytr,'-or'); hold on plot(YtrNet,'-sb'); hold off
figure(2) plot(Yts,'-or'); hold on plot(YtsNet,'-sb'); hold off
figure(3) t = -1:.1:1; plot(t,t,'b','linewidth',2) hold on plot(Ytr,YtrNet,'ok') hold off
figure(4) t = -1:.1:1; plot(t,t,'b','linewidth',2) hold on plot(Yts,YtsNet,'ok') hold off