

Article

# Application of Artificial Intelligence for Predicting CO<sub>2</sub> Emission Using Weighted Multi-Task Learning

Mohammad Talaei <sup>1</sup>, Majid Astaneh <sup>2</sup>, Elmira Ghiasabadi Farahani <sup>1</sup> and Farzin Golzar <sup>3,\*</sup>

<sup>1</sup> Department of Energy Engineering, Sharif University of Technology, Tehran 1458889694, Iran; mohammad.talaei@sharif.edu (M.T.); elmira.farahani@energy.sharif.edu (E.G.F.)

<sup>2</sup> Northvolt Battery Systems AB, Alströmergatan 20, 112 47 Stockholm, Sweden; majid.astaneh@northvolt.com

<sup>3</sup> Department of Energy Technology, KTH-Royal Institute of Technology, 114 28 Stockholm, Sweden

\* Correspondence: fargo@kth.se

**Abstract:** Carbon emissions significantly contribute to global warming, amplifying the occurrence of extreme weather events and negatively impacting the overall environmental transformation. In line with the global commitment to combat climate change through the Paris Agreement (COP21), the European Union (EU) has formulated strategies aimed at achieving climate neutrality by 2050. To achieve this goal, EU member states focus on developing long-term national strategies (NLTSS) and implementing local plans to reduce greenhouse gas (GHG) emissions in alignment with EU objectives. This study focuses on the case of Sweden and aims to introduce a comprehensive data-driven framework that predicts CO<sub>2</sub> emissions by using a diverse range of input features. Considering the scarcity of data points, we present a refined variation of multi-task learning (MTL) called weighted multi-task learning (WMTL). The findings demonstrate the superior performance of the WMTL model in terms of accuracy, robustness, and computation cost of training compared to both the basic model and MTL model. The WMTL model achieved an average mean squared error (MSE) of 0.12 across folds, thus outperforming the MTL model's 0.15 MSE and the basic model's 0.21 MSE. Furthermore, the computational cost of training the new model is only 20% of the cost required by the other two models. The findings from the interpretation of the WMTL model indicate that it is a promising tool for developing data-driven decision-support tools to identify strategic actions with substantial impacts on the mitigation of CO<sub>2</sub> emissions.

**Keywords:** CO<sub>2</sub> emissions prediction; artificial intelligence; weighted multi-task learning



**Citation:** Talaei, M.; Astaneh, M.; Ghiasabadi Farahani, E.; Golzar, F. Application of Artificial Intelligence for Predicting CO<sub>2</sub> Emission Using Weighted Multi-Task Learning. *Energies* **2023**, *16*, 5956. <https://doi.org/10.3390/en16165956>

Academic Editor: Kun Mo Lee

Received: 30 June 2023

Revised: 30 July 2023

Accepted: 4 August 2023

Published: 12 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Global warming has a meaningful impact on the transformation of the environment, which strengthens the occurrence rate of extreme weather events in the long term. The Eurobarometer report on climate change points out that more than 90% of the European Union (EU) citizens consider climate change to be a serious problem [1]. In alignment with the global climate action commitment to the Paris Agreement (COP21), the EU has developed strategies towards the climate-neutrality objective by 2050 [2]. For this aim to be achieved, the EU member states aim at developing national long-term strategies (NLTSS) via the development and implementation of local plans to reduce the greenhouse gas (GHG) emissions in line with EU objectives. Salvia et al. [3] provided a comprehensive review on the climate action plans of 327 cities in the EU. The authors aimed at investigating the existing correlation between the level of climate-mitigation ambitions and several factors, such as the city size and its regional location, the type of the plan, and membership of the climate networks. It was concluded that it is required that the cities urgently double their ambitions to meet the climate action criteria. Moreover, it was further highlighted that the national-level plans and legislation impact the local targets to achieve EU objectives. Among EU members, Sweden has set the overall goal of climate neutrality by 2045, even

though the strategy is mainly outlined in a qualitative context [4]. To this end, the Swedish government states that the climate policy framework for the country should be founded on sectoral long-term and time-based emissions targets [5]. To address the aforesaid challenge, this work, by considering the study case of Sweden, aims at presenting a comprehensive data-driven framework to predict CO<sub>2</sub> emissions as a function of a diverse set of input features. This, in turn, provides time-based NLTS to facilitate the national-level policy-making process in alignment with EU targets.

#### *Background and Research Gap*

Short- and long-term carbon emission prediction and its consequence on climate change have gained ever-increasing attention in technological, policymaking, and socio-economic studies in recent years [6]. Sognaes et al. [7] presented a multi-model analysis approach to predict CO<sub>2</sub> emissions by comparing technologically rich models (e.g., GCAM, TIAM, and MUSE) and macro-economic approaches (e.g., ICES, GEMINI, and E3ME). Their study showed that the precision of the outcomes was highly dependent on the accuracy level of the baseline emissions, as well as current policies (CPs) and nationally determined contributions (NDCs). This reflects that the development of individual national-level models is of crucial importance in the reliable prediction of CO<sub>2</sub> emissions. Mansfield et al. [8], by highlighting the computational complexity of the climate models, proposed a machine-learning-based method to capture climate change trends from short-term simulations. Their study focused on temperature changes as the climate change response factor rather than the carbon emission. However, the presented methodology reveals the pivotal role of data-driven approaches in accelerating environmental projections by uncovering the early contributors of long-term responses.

Artificial neural networks (ANNs) are categorized as modern artificial intelligence models for being applied in a wide variety of non-linear problems, which comprise predictions, projections, simulations, pattern identification, scheduling, and optimization [9–12]. Due to the rise and progress of Industry 4.0, modern energy systems have acknowledged the vast potential of artificial intelligence. As a result, they are now envisioning the utilization of more complex, creative algorithms across various applications [13]. Through their investigation of various forecasting models, Hu et al. [14] suggested employing deep-learning, data-driven, and intelligent algorithms as innovative approaches to develop prediction models for energy consumption and carbon emissions. Klyuev et al. presented a comprehensive review of different classical and modern methods of forecasting electricity consumption. The authors highlighted the importance of classifying the forecast method based on the forecasting horizon to distinguish between short- to long-term predictions [15]. Aryai et al. [16] applied a machine-learning method for day-ahead predicting of the emissions' intensity for the Australian National Electricity Market. Authors concluded that the extremely randomized trees regressor model they proposed outperformed other classic machine-learning algorithms (Extreme Learning Machine (ELM), multilayer perceptron (MLP), and decision tree (DT)). The ANNs have also gained enormous popularity in national-level energy planning, financial assessment, and environmental forecasting studies in recent years. For instance, Kermanshahi et al. [17] implemented ANN techniques based on ten input factors to investigate current and future trends of electric loads in Japan. Azadeh et al. [18] developed an integrated, flexible model, using ANN and computer simulations to predict Iran's national electrical energy consumption. It was shown that the proposed algorithm could provide a dynamic structure for the forecasting. Sadri et al. [19] developed an ANN-based energy and environmental planning framework using historical data for the transportation sector in developing countries where the required data are either unavailable or limited. Mason et al. [20] used evolutionary neural networks to predict short-term CO<sub>2</sub> emissions, demanded power, and wind energy generation in Ireland. Marjanovic et al. [21] aimed at predicting the gross domestic product (GDP) based on CO<sub>2</sub> emissions, using an extreme learning method. Sun and Xu [22] applied ANNs optimized by a genetic algorithm to evaluate the financial security of the electric power industry in

China. The study's results showed superior accuracy and convergence speed advantage compared to the least-square support vector machine and traditional backpropagation neural network methods. Sahraei et al. [23] introduced a novel hybrid metaheuristic and ANN method for predicting the energy demand of the transportation sector. The authors optimized the coefficients for energy demand prediction based on different input features, such as GDP, vehicle-km, population, and oil price in Turkey. The outcomes revealed that the integration of ANN with Particle Swarm Optimization (ANN-PSO) outperforms other combinations.

In the above-referenced studies, the main focus was on the energy sector or economic growth and security. Therefore, CO<sub>2</sub> emission was considered either an input feature or a less prioritized output.

Predicting CO<sub>2</sub> emissions is categorized as a supervised machine-learning problem. The selection of a supervised machine-learning algorithm depends on the nature of the data and the specific problem being investigated. Tree-based models blend the predictions of various decision trees. The decision trees learn by employing interpretable decision rules based on the information theory [24]. They are fast to train, require relatively little data for satisfactory performance, and are less prone to overfitting. ANNs, on the other hand, learn through an error backpropagation process. ANNs are well-suited for handling complex non-linear relationships. They can automatically learn hierarchical representations from raw data, enabling them to capture intricate patterns. ANNs and deep-learning models are known as the best data-driven models for image and text data [25,26]. When dealing with tabular data, employing ensemble techniques and tree-based models such as Random Forest and XGBoost mostly yields more precise outcomes [27]. Nevertheless, tree-based models face limitations when it comes to predicting future outcomes in regression problems. In fact, when the range of target outputs has the potential to change, tree-based models are unable to predict values that fall outside the training range. To be more specific, since the tree-based model's predictions are based on the mean value of the training samples in each leaf node, the predictions will never be outside the range of the target variable observed in the training data. When applying these models to real-world scenarios, this aspect is essential to consider [24]. In situations where, for example, the minimum CO<sub>2</sub> emission during the training phase surpasses the expected CO<sub>2</sub> emission for the next year, tree-based models are unable to accurately forecast such a scenario.

Conversely, ANNs do not face this constraint. Since most developed countries strive to reduce their national-level CO<sub>2</sub> emissions over time, employing ANN-based models is a suitable approach for developing frameworks to predict carbon emissions.

A review of the literature revealed that, in the context of ANN-based CO<sub>2</sub> emission prediction in energy systems, the vast majority of studies has been either focused on a specific industry or an energy sector [28–32]. For instance, Safa et al. [29] aimed at estimating CO<sub>2</sub> emissions from wheat production farms in New Zealand, using ANN and multiple linear regression (MLR) models. The authors concluded that the ANN predictions had a better performance compared to the MLR outcomes, with an approximately 37% lower root mean square error (RMSE). Singh et al. [28] proposed a deep-learning modeling approach to predict CO<sub>2</sub> emissions from road transport vehicles. The results showed that the long short-term memory (LSTM) model based on recurrent neural networks (RNNs) performed remarkably better than other models. Several studies have concentrated on the global CO<sub>2</sub> emissions at the national, regional, or EU level. Ma et al. [33] developed a machine-learning-based algorithm, using the Gaussian process regression, to analyze CO<sub>2</sub> emissions in China. Five independent variables, namely economic growth, energy consumption, population, industrialization, and income, were considered in this study. It was shown that the introduced method provided more accurate predictions compared to traditional least-squares and the robust least-squares models. Du et al. [34] introduced an improved backpropagation neural network and genetic algorithm (BP-GA) framework to predict mid- and long-term CO<sub>2</sub> emissions in Jiangsu Province. The authors highlighted that the proposed method in their study led to higher prediction accuracy than other

conventional methods, with a maximum relative error of 0.76%. Dozic et al. [9] used the ANN approach for forecasting CO<sub>2</sub> emissions and testing EU long-term energy policy targets. They determined that ANN with a cascade forward/backpropagation structure yielded reasonable accuracy for this purpose. Recently, Brito et al. [35] combined a scenario analysis approach with ANN to quantitatively correlate CO<sub>2</sub> emissions, energy matrix, and burning in Brazilian biomes. The authors used the developed model to define different scenarios that help Brazil balance development and sustainability through CO<sub>2</sub> prediction and control. Han et al. [36] introduced improved residual neural networks to predict carbon emissions in 24 different countries/regions based on the needs for various primary energy resources from 2009 to 2020. According to the authors' investigation, the CO<sub>2</sub> emissions of Russia, the United States, China, India, and Japan exceeded 1000 Mt in 2020. Moreover, the CO<sub>2</sub> emissions of Brazil, Germany, South Africa, and South Korea exceeded 400 Mt. This is while the CO<sub>2</sub> emissions of Sweden were less than 100 Mt. However, the authors noted that Sweden needs to precisely handle the relationship between environmental protection and economic development by using wind and solar power instead of primary energy to reduce future CO<sub>2</sub> emissions.

Various sectors and processes contribute to global CO<sub>2</sub> emissions. Accordingly, a wide variety of individual policies and actions need to be employed to decarbonize the energy system. None of the previous studies has focused on predicting sector-by-sector CO<sub>2</sub> emissions; therefore, their outcomes are often qualitative and vague regarding decision-making at the sectoral level to cover climate action policies. The present work aims to fill the aforementioned knowledge gap in alignment with the Swedish government's strategy for identifying national-level sectoral emission targets. To this end, this research was initially inspired by the multi-task learning (MTL) concept introduced by Zhang and Yang [37] for different engineering and natural science applications that contain multiple outputs. While having sufficient data is crucial to develop reliable machine-learning models, only a few records of data points are available in this work since CO<sub>2</sub> emissions were once a year. The objective of this work is to utilize artificial intelligence and data-driven methods to predict the CO<sub>2</sub> emissions of Sweden at the national level, using available historical data from 1990 to 2019. We present an improved version of MTL named weighted multi-task learning (WMTL), which is capable of obtaining the most information out of the available limited data where the output tasks can be weighted based on their prioritization level. The proposed approach assigns weights to output tasks and selects pertinent subtasks by using a devised algorithm. To this aim, Bayesian optimization is employed to optimize the hyperparameters of the models.

## 2. Materials and Methods

While it is an excellent way to measure CO<sub>2</sub> emission with emission factors, it would not be an appropriate approach to predict long-term CO<sub>2</sub> emissions. Since the emission factors are highly dependent on the technology, this approach is blind to relationships between sectors, the big picture, and the megatrends that evolve the technology. Furthermore, considering the rapid pace of technological advancements in today's industries, relying on emission factors to predict medium-term CO<sub>2</sub> emissions is not viable. Additionally, data pertaining to various technologies and their corresponding emission factors are currently unavailable.

On the other hand, data-driven methods can appropriately find the non-linear relationships between input features and outputs. Thus, regardless of the technologies impacting CO<sub>2</sub> emissions, we can rely on the data and machine-learning methods, as long as a well-developed model is presented.

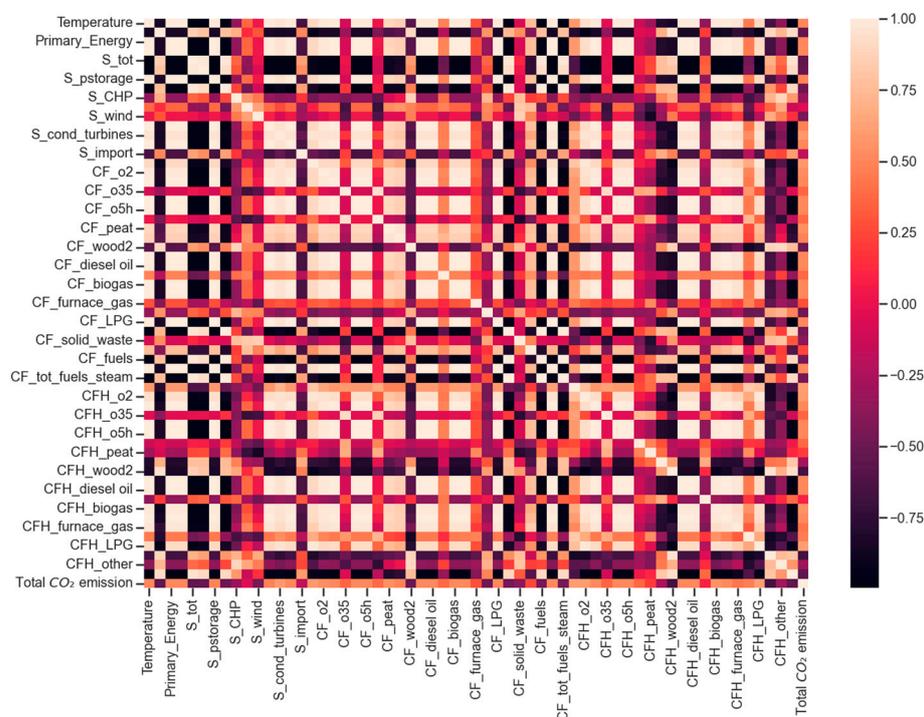
Considering the complexities associated with different sectors and industries, investigating all relationships among impactful parameters and CO<sub>2</sub> emissions with analytical methodologies is impossible. Data-driven models have been proven to be implemented as an alternative to analytical approaches due to no need for detailed knowledge of internal

process parameters. These techniques greatly reflect the complex dependences of a given process, without considering detailed mechanisms of specified processes.

### 2.1. Input Data

The first step in developing data-driven models is the identification of impactful input parameters to understand how these parameters affect the outputs. Dozic et. al. [9] concluded that the gross domestic product (GDP); population; average annual air temperature; Total Primary Energy Supply (TPES); electricity consumption; and share of renewable, nuclear, natural gas, total petroleum products, and solid fuels energy in the TPES are the crucial parameters for forecasting CO<sub>2</sub> emissions in the European Union. The current study follows similar parameters with more details for more complex parameters. For example, the electricity supply is divided into main resources, such as hydro, pumped storage, nuclear, main activity producer CHP, autoproducer CHP, wind, solar, condensing turbines, gas turbines for reserve and others, and import. Moreover, the types of fuels used in different types of power plants for electricity production are considered as separate input parameters. In addition, the consumptions of different fuels for steam and hot water production by type of power plants are identified as important input parameters. All in all, 61 parameters are considered as the input parameters for the data-driven model to identify their impacts on Sweden's total CO<sub>2</sub> emission and emissions from different sectors, such as agriculture, domestic transport, international transport, off-road vehicles, electricity and district heating, household heating, industry, land use, solvent use and other product, and waste management. The historical data of these parameters from 1990 to 2019 are taken into account. Table A1 in Appendix A summarizes all input and output parameters.

Figure 1 illustrates the correlation heatmap of our dataset. As depicted in the figure, it is evident that nearly all features exhibit a correlation with the target value, which is the total CO<sub>2</sub> emissions.



**Figure 1.** Correlation heatmap of the dataset (Only some non-overlapping labels are shown).

### 2.2. Preprocessing Input Features

ANNs are influenced by the magnitude of various input features. This occurs because features with larger ranges and magnitudes can disproportionately affect the output compared to other features. To address this, it is important to normalize all input features to

a consistent scale. Min–max scaling is a technique that is commonly used to achieve this. The process involves subtracting the minimum value of the feature and then dividing it by the range (maximum value minus minimum value). This normalization brings the values of the feature between 0 and 1.

However, the existence of outliers in the data, which are extreme values that deviate substantially from the majority of data points, can distort the minimum and maximum values employed for scaling. As a result, the scaling operation can lead to a compression of the majority of the data points into a narrow range, while the outliers may occupy a disproportionate part of the range. On the other hand, outliers do not affect standard scaling, making it a perfect choice for this dataset, where many features have extreme values. Standard scaling is defined by Equation (1):

$$X' = \frac{X - \mu}{\sigma} \quad (1)$$

where  $X'$  is the scaled feature,  $X$  is the raw feature,  $\mu$  is the mean of the raw feature, and  $\sigma$  is the standard deviation of the raw feature.

### 2.3. Model

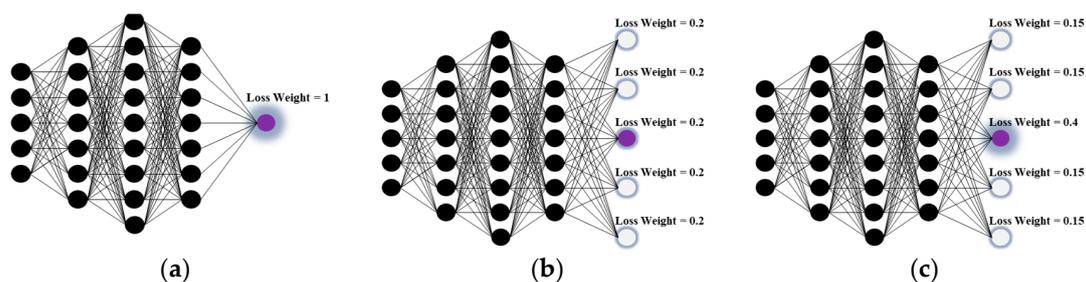
Figure 2 illustrates the schematic representations of the three models assessed in this study, all of which are categorized as ANN architectures. Equations (2)–(4) show the forward-propagation formulation between layers. Essentially, this process entails a sequence of matrix multiplications between layers, followed by the application of non-linear functions known as activation functions.

$$Z^{[l]} = w^{[l]}a^{[l-1]} + b^{[l]} \quad (2)$$

$$a^{[l]} = g^{[l]}(Z^{[l]}), \quad (3)$$

$$a^{[l]} = x, \quad (4)$$

where  $a^{[l]}$  is the output of layer  $l$ ,  $Z^{[l]}$  is the result of applying a linear function to the input of layer  $l$ ,  $w^{[l]}$  represents the weights between layer  $l$  and layer  $l-1$ ,  $b^{[l]}$  is the bias of layer  $l$ ,  $g^{[l]}$  is the activation function of layer  $l$ , and  $x$  represents the input features.



**Figure 2.** Schematic representation of models in the study: (a) basic model, (b) MTL model, and (c) WMTL model (The main-task is represented by the purple color. The T values in the figure are purely illustrative examples).

Equations (5) and (6) outline the formulation for the backward-propagation process, which is the learning process responsible for updating the parameters of the model (weights and biases):

$$w^{[l]} := w^{[l]} - \alpha dw^{[l]} \quad (5)$$

$$b^{[l]} := b^{[l]} - \alpha db^{[l]}, \quad (6)$$

where  $dw^{[l]}$  is the gradient of the model error with respect to  $w^{[l]}$ ,  $db^{[l]}$  is the gradient of the model error with respect to  $b^{[l]}$ , and  $\alpha$  is the learning rate.

The gradients are computed through a procedure known as error backpropagation, which entails calculating partial derivatives by using the chain rule. The learning rate is a hyperparameter of the model that requires tuning.

The basic model is a single-task model focused solely on predicting the main task, which, in this context, corresponds to the total CO<sub>2</sub> emission of Sweden. The MTL model, on the other hand, is a multi-task learning model that includes subtasks. Lastly, the WMTL model is a weighted multi-task learning model that incorporates customized loss weights for the output variables. The objective of this study is to achieve high performance specifically for the main task, which is why the loss associated with the main task is relatively higher in the WMTL model. It is important to note that the provided values in Figure 2 are purely illustrative examples.

### 2.3.1. Basic Model

In this work, the basic model refers to the conventional ANN architecture with only one target output in its last layer (Figure 2a). This target is the main target to be predicted, which is the total CO<sub>2</sub> emissions of Sweden in the given year. The total loss of the basic model is calculated by the mean squared error (MSE) function, which is represented in Equation (7):

$$TotalLoss = \frac{1}{m} \sum_{j=1}^m (Y_j^p - Y_j^a)^2, \quad (7)$$

where  $Y^p$  is the predicted value of the main task,  $Y^a$  is the actual value of the main task, and  $m$  is the number of data points.

### 2.3.2. Multi-Task Learning Model

Multi-task learning is a supervised neural network architecture that has multiple target outputs (Figure 2b). The main advantage of multi-task learning is that, with a careful design, the prediction accuracy of the main task would be higher in comparison with a single-task model [37]. In fact, the presence of other subtasks contributes to the generation of valuable features that can enhance the accuracy of predictions. Equations (8) and (9) present the calculation of the overall loss for the MTL model, and Equation (10) depicts the formulation of the loss function for each individual task, using the mean squared error:

$$TotalLoss = \sum_{i=1}^m W_i loss_i, \quad (8)$$

$$W_i = \frac{1}{n}, \quad (9)$$

$$loss_i = \frac{1}{m} \sum_{j=1}^m (Y_{ij}^p - Y_{ij}^a)^2, \quad (10)$$

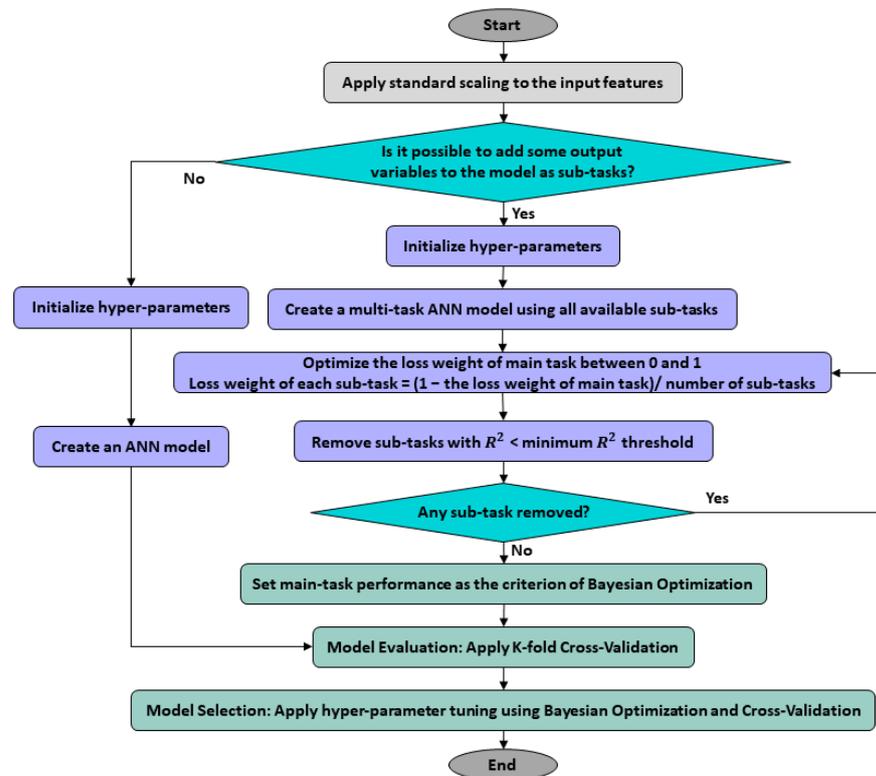
where  $W_i$  is the weight of the  $i$ -th task,  $loss_i$  is the MSE of the  $i$ -th task, and  $n$  is the number of tasks.

### 2.3.3. Weighted Multi-Task Learning Model

The downside of the MTL model is that each subtask's loss weight is equal to that of the main task. In fact, the accumulated loss weight fraction of all subtasks would be higher than the loss weight fraction of the main task, which, in turn, prevents the main task from contributing enough to the loss function. Furthermore, not all subtasks are necessarily helpful in increasing the main task accuracy. To address these problems, a WMTL learning model was developed in this study.

Figure 3 illustrates the algorithm utilized for designing the WMTL model. Once the raw input features are transformed, an assessment is conducted to determine the feasibility of developing a WMTL model by adding subtasks. In the case of the WMTL model, once

the hyperparameters are initialized, the first step involves constructing a multi-task model that incorporates all the accessible subtasks alongside the main task.



**Figure 3.** Algorithm for developing a WMTL model.

One distinction between the MTL and WMTL models is that, in the case of the WMTL model, the loss weight assigned to the main task is optimized using the coefficient of determination ( $R^2$ ) as the criterion [38]. As a result, the loss weights of the subtasks are subsequently updated based on the algorithm's specifications.

During the development of multi-task models, we observed that certain subtasks which exhibited poor prediction performance negatively impacted the main task's predictability. In fact, our observations revealed that by eliminating the subtasks with an  $R^2$  below a minimum threshold of 0.65, the predictability of the main task demonstrated notable improvement. This observation was consistent for both the MTL and WMTL models.

$$W_{sub-task} = \frac{1 - W_{main-task}}{N_{sub-tasks}} \quad (11)$$

#### 2.4. Hyperparameter Tuning and Bayesian Optimization

In order to achieve a satisfactory performance for each of the three models, a comprehensive search space of hyperparameters was defined. These hyperparameters are presented in Table 1. Considering the size of the search space, which is almost 50,000, it is not wise to use the Grid Search method for the tuning. In Grid Search, all parameter sets in the search space are evaluated. The Random Search method is not efficient, either, because it simply evaluates a random subset of the search space. The advantage of the Bayesian optimization is that it does not exhaustively search the entire search space to reach the optimal value. Instead, it intelligently explores a subset of the search space based on its observations and previous evaluations. It employs a probabilistic model, such as Gaussian processes, to model the objective function and uses an acquisition function to guide the search toward promising regions. By iteratively evaluating and updating the

model, Bayesian optimization focuses on areas with higher potential for finding the optimal solution, making it an efficient approach for hyperparameter tuning.

**Table 1.** Search space of the hyperparameters for the models.

Hyperparameter	Values
Learning rate	[0.05, 0.01, 0.005, 0.001, 0.0001, 0.00001]
No. of hidden layers	[1, 2, 3, 4, 5]
No. of neurons in each layer	[32, 64, 128, 256, 512, 1024]
No. of epochs	[1, 5, 10, 20, 100, 200]
Activation function of each layer	[linear, sigmoid, tanh, relu, selu]
Batch size	[2, 4, 8]
Dropout value	[0, 0.2, 0.4]

The objective of the Bayesian optimization is to choose the optimal hyperparameters that result in the lowest validation error. In this work, the search space for potential hyperparameters, denoted as  $x$ , corresponds to the table shown in Table 1. The objective function, represented by  $f$ , aims to minimize the cross-validation error, specifically using the MSE criterion in this case. The optimization process in the Bayesian optimization algorithm (BOA) can be described as shown below:

$$x^+ = \arg \arg \max_{x \in A} f(x), \quad (12)$$

In this context, the symbol  $A$  represents the search space of  $x$ . Bayesian optimization is derived from Bayes' theorem [39]. It states that the posterior probability,  $P(M|E)$ , of model  $M$ , given evidence data  $E$ , is proportional to the likelihood  $P(E|M)$  of observing  $E$  given model  $M$ , multiplied by the prior probability,  $P(M)$ :

$$P(M|E) \propto P(E|M)P(M), \quad (13)$$

Bayesian optimization employs a probabilistic model, such as Gaussian processes, to model the objective function. A Gaussian process is a type of random process where any finite subset of random variables follows a multivariate Gaussian distribution. This process assumes that similar inputs yield similar outputs, thereby establishing a statistical model of the function [40].

Once the posterior distribution of the objective function is obtained, Bayesian optimization employs the acquisition function to determine the maximum value of the function,  $f$ . Typically, it is assumed that a higher value of the acquisition function corresponds to a larger value of the objective function,  $f$ . Therefore, maximizing the acquisition function is equivalent to maximizing the function,  $f$ .

$$x^+ = \arg \arg \max_{x \in A} (U(D)), \quad (14)$$

where  $x^+$  represents the position where the function,  $f$ , is maximized after obtaining  $t$  sample points.

### 2.5. Model Interpretation

When a machine-learning model is trained to be a decision-making tool, feature importance methods are utilized to obtain the effect of each feature on the model's outputs. The permutation feature importance method was developed for such a purpose. In order to obtain the importance of feature  $F1$  with respect to the output,  $F1$  will be randomly shuffled in the dataset, and then, after making new predictions using the trained model, the level at which the accuracy of prediction is decreased shows the importance [41]. In this work, the permutation importance and partial dependency of each feature are calculated with respect to the main task, which can help policymakers prioritize the factors for reducing CO<sub>2</sub> emissions.

Although permutation feature importance can calculate the effect of each feature on the model's outputs, it does not provide any information about the direction in which each feature is related to the output. However, partial dependency-based methods [42] are more informative about the direction. In this paper, the direction for each feature is obtained using partial dependency. Thus, the model interpretation method in this paper is a combination of permutation feature importance and partial dependency-based methods.

### 3. Results

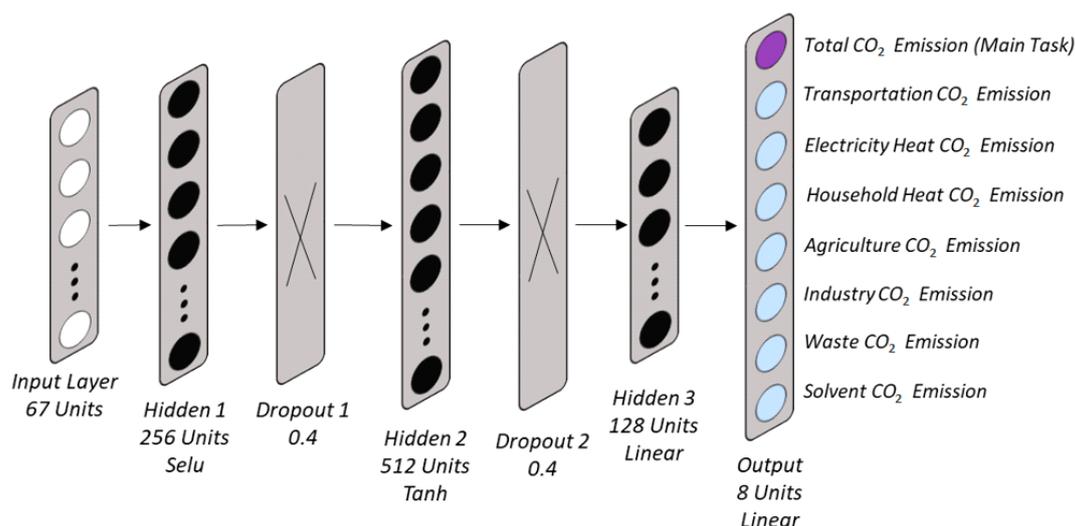
#### 3.1. Developed Models Description

The hyperparameters for all three model types were optimized. Table 2 presents the optimal hyperparameters for each model. Interestingly, the number of epochs indicates that training a WMTL model requires only one-fifth of the computation cost compared to the other two models. All models have three hidden layers.

**Table 2.** Optimal hyperparameters of the models.

Model	No. of Hidden Layers	No. of Epochs	Learning Rate	Batch Size	Activation Function and No. of Neurons
Basic Model	3	100	0.01	8	128 tanh, 512 tanh, 16 linear, 1 linear
MTL	3	100	0.001	2	256 selu, 512 tanh, 128 linear, 8 linear
WMTL	3	20	0.001	4	256 selu, 512 tanh, 64 linear, 8 linear

Figure 4 is a representation of the tuned WMTL model. Following the implementation of the aforementioned algorithm for the WMTL model, a total of seven subtasks were retained within the model. The subtasks can be identified in the figure. The optimal loss weight assigned to the main task was 0.51, while all other tasks were assigned a loss weight of 0.07.



**Figure 4.** Representation of the tuned WMTL model.

When working with machine-learning models, a critical consideration is to avoid overfitting. Overfitting occurs when a model becomes too complex and starts to memorize the training data rather than learning general patterns that can be applied to unseen data. In order to prevent overfitting, we monitored the training and test losses during the training epochs, enabling us to stop the training once the test loss starts to increase or plateau while the training loss continues to decrease (see Figure 5). We also regularized our models by

employing two dropout layers with a dropout rate of 0.4, which can be seen in Figure 4. A dropout layer randomly deactivates a proportion of its previous layer's neurons in each epoch of the training phase. This ensures that individual neurons do not excessively rely on one another and minimizes co-adaptation among them. As a result, the model becomes more robust and less likely to overfit the training data.

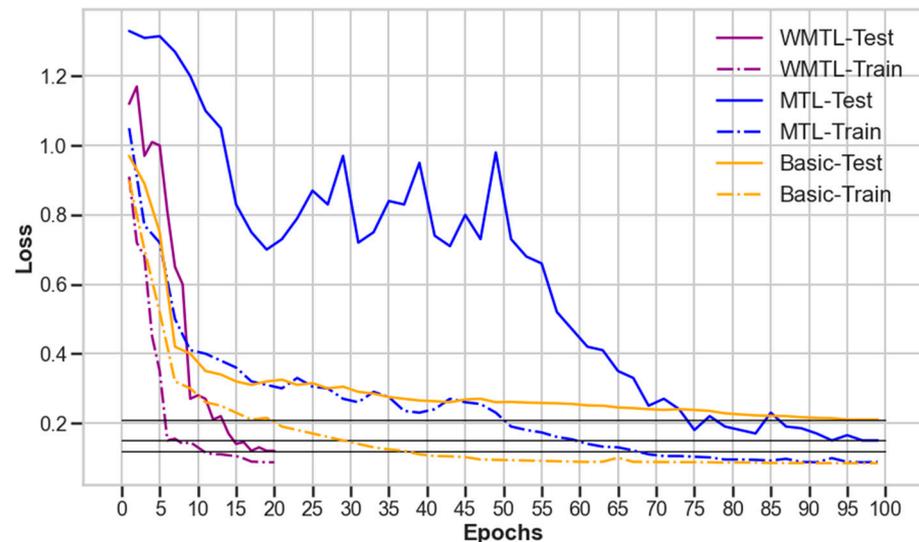


Figure 5. Training and test loss over epochs of training.

The SELU activation function, short for “scaled exponential linear unit”, was initially introduced to facilitate the creation of high-level abstract representations in shallow networks, such as the one employed in this study. Additionally, it enhances the robustness of the learning process [43]. Given these advantages, the SELU activation function was incorporated into the search space for activation functions, and, consequently, it was unsurprisingly selected as the activation function for the first hidden layer in the Bayesian optimization process.

### 3.2. Testing and Validation of the Developed Models

To obtain more accurate evaluations of the models and guard against overfitting, K-fold cross-validation was employed. This technique involves dividing the data into six equal subgroups. One subgroup is designated as the test set, while the remaining subsets serve as the training set. This process is repeated, ensuring that each subgroup is used exactly once for testing. By doing so, overfitting a specific test set is avoided, and the risk of reporting inflated evaluation results is mitigated.

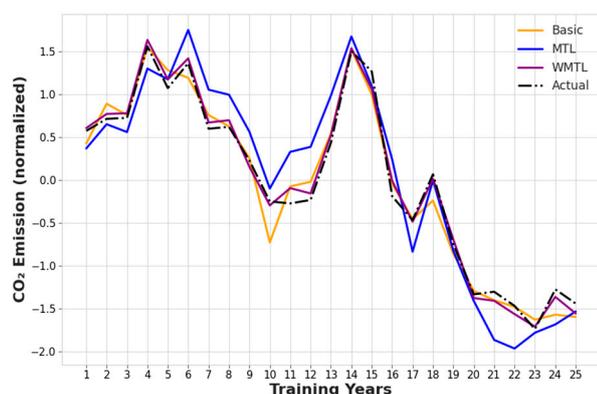
Figure 5 displays the training and test loss over the training epochs for the optimized versions (after hyperparameter tuning) of the models presented. The horizontal black lines represent the converged loss of the models. The figure shows that the WMTL model not only reached a lower level of loss, but it also converged faster compared to the other models. Additionally, the MTL model performed better than the basic model. However, during training, the test loss of the MTL model fluctuated, which can be attributed to the model's attempt to equally reduce the loss of subtasks.

Table 3 demonstrates that the WMTL model exhibited a superior performance compared to both the basic model and the MTL model. The average mean squared error (MSE) of the WMTL model across folds was 0.12, which is lower than the MTL model's MSE of 0.15 and the basic model's MSE of 0.21. The robustness of a model can be assessed by examining its worst prediction across the folds. Interestingly, the WMTL model also demonstrates the highest level of robustness among the models evaluated.

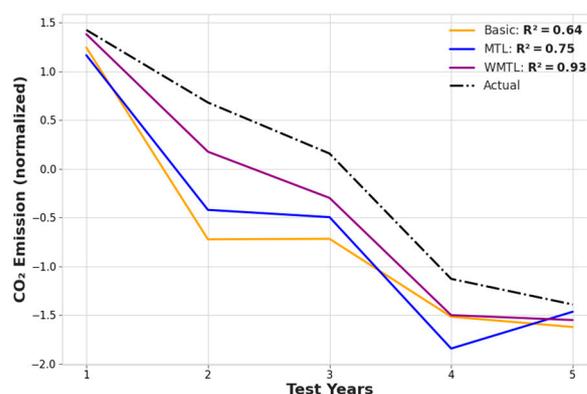
**Table 3.** Performance of best single-task, best multi-task, and best weighted multi-task models on test data.

	Basic Model		MTL		WMTL	
	R <sup>2</sup>	MSE	R <sup>2</sup>	MSE	R <sup>2</sup>	MSE
Fold 1	0.80	0.19	0.89	0.10	0.81	0.17
Fold 2	0.79	0.15	0.87	0.09	0.85	0.10
Fold 3	0.70	0.22	0.84	0.11	0.84	0.11
Fold 4	0.64	0.40	0.75	0.27	0.93	0.08
Fold 5	0.78	0.18	0.83	0.13	0.80	0.18
Fold 6	0.93	0.10	0.88	0.18	0.93	0.10
Mean	0.77	0.21	0.84	0.15	<b>0.86</b>	<b>0.12</b>
Worst	0.64	0.40	0.75	0.27	<b>0.80</b>	<b>0.18</b>
Best	0.93	0.10	0.89	0.09	<b>0.93</b>	<b>0.08</b>

Figure 6 illustrates the predictions of the training and test sets for the models using fold 4. It is evident that the WMTL model achieves more accurate predictions for both the training and test sets.



(a)



(b)

**Figure 6.** Illustration of predicted values of the fold 4 of models: (a) training years and (b) test years.

### 3.3. Interpretation and Feature Importance

Figure 7 displays the ordered outcomes of the most significant features determined through permutation feature importance. However, it is essential to note that permutation feature importance does not indicate the direction of the relationship between the features and the output. In this study, the selected WMTL model is interpreted using a combination of Permutation Feature Importance and Partial Dependency analysis to uncover the directionality of these relationships, as depicted in Figure 8. Specifically, in our approach, once we establish the importance of features, we can determine whether they contribute to the mitigation of CO<sub>2</sub> emissions. These results can serve as a foundation for devising effective policies and making informed decisions regarding various aspects of energy supply and consumption. However, it is crucial to note that these interpretations do not necessarily establish causal relationships between input features and the total CO<sub>2</sub> emission. In this regard, these interpretations aid in the decision-making process and expert assessments, making them valuable primary tools for experts and policymakers.

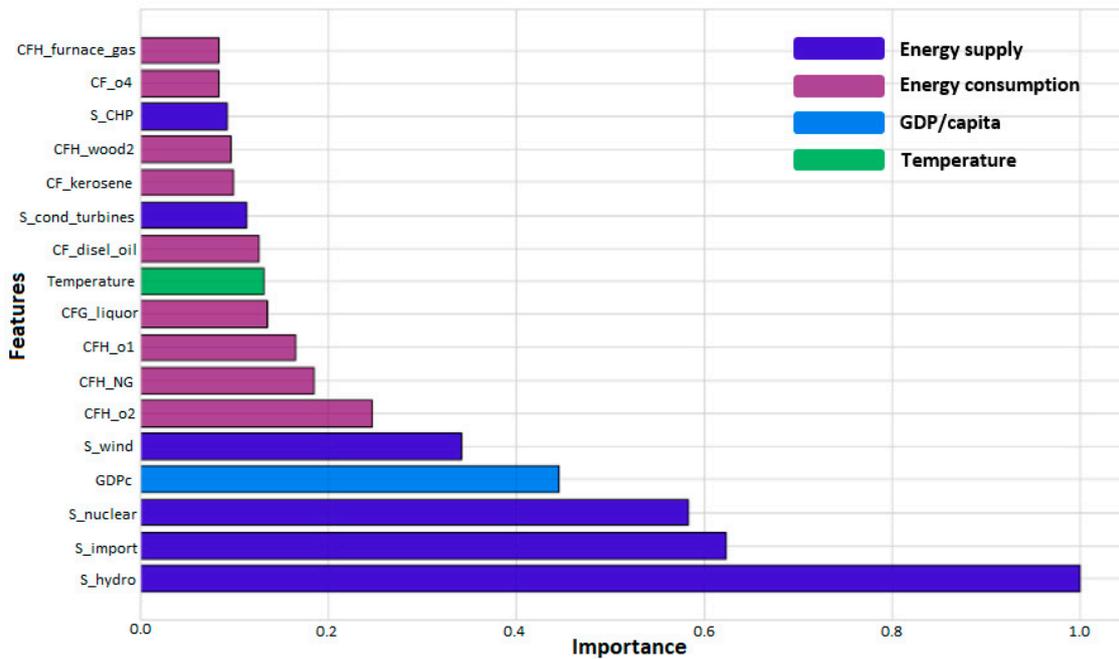


Figure 7. Feature importance of top important features of the WMTL model (The full names and statistical descriptions of all parameters can be found in Appendix A Table A1).

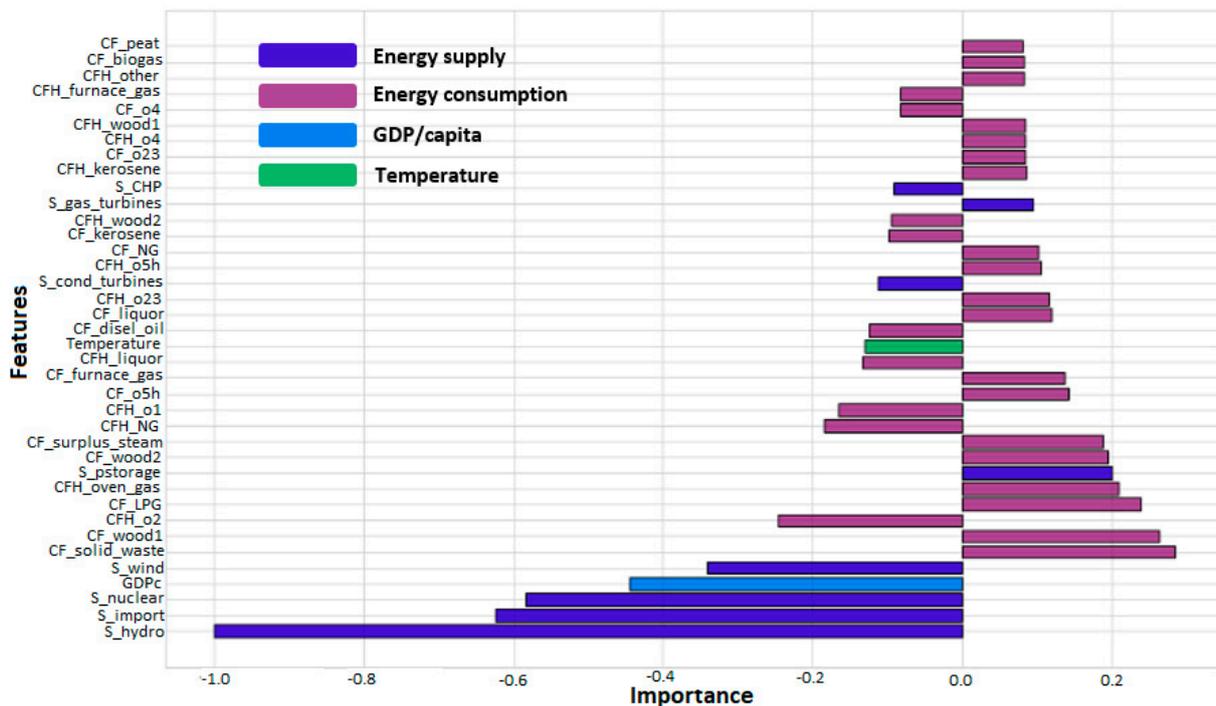


Figure 8. Feature importance with direction of top important features of the WMTL model. (The full names and statistical descriptions of all parameters can be found in Appendix A Table A1.)

Based on the findings of the directed feature importance analysis presented in Figure 8, it is indicated that hydropower energy supply has the greatest impact on the mitigation of CO<sub>2</sub> emissions. Additionally, importing energy, nuclear energy, GDP/capita, and wind energy are also identified as influential factors in reducing CO<sub>2</sub> emissions. Nevertheless, there appears to be no causal relationship between GDP per capita and CO<sub>2</sub> emissions. The insights and correlations presented can be comprehended by analyzing Figure 7. In this graphical representation, the size of each bar corresponds directly to the extent of impact

that a particular parameter would have on the CO<sub>2</sub> emissions. Therefore, the larger the bar in the graph, the more significant the influence of that specific parameter on the overall CO<sub>2</sub> emission levels. Despite its informative nature, Figure 7 lacks the visualization of whether each parameter has a positive or negative impact on CO<sub>2</sub> emissions. To address this, Figure 8 is introduced, specifically designed to determine the direction of each parameter's impact. By carefully examining Figure 8, it becomes evident that negative parameters signify that an increase in their value would lead to a decrease in CO<sub>2</sub> emissions, while positive parameters indicate the opposite, resulting in an increase in CO<sub>2</sub> emissions. This comprehensive representation enables a clearer understanding of the relationships between various parameters and their respective impacts on CO<sub>2</sub> emission levels.

#### 4. Discussion

In this section, we support the outcomes of the proposed model in this work by providing key findings from the recently published research in the literature. The results of this research for the case study of Sweden indicated that hydropower energy supply has the greatest impact regarding the mitigation of CO<sub>2</sub> emissions. It was also shown that importing energy, nuclear energy, GDP/capita, and wind energy are also identified as influential factors in reducing CO<sub>2</sub> emissions. Berga [44] emphasized the importance of hydropower in global climate-change mitigation and adaptation, as it prevents about 9% of global annual CO<sub>2</sub> emissions. By referring to the IRENA REMAP 2030 scenario, the authors mentioned that doubling the global share of renewable energy requires 2200 GW of global hydropower capacity to achieve its targets. Muhsin et al. [45] studied the impact of hydropower energy to reduce CO<sub>2</sub> emissions in European Union countries. According to their correlation analyses, hydropower and CO<sub>2</sub> are negatively and remarkably connected with each other in most of the investigated countries, and Sweden held the greatest correlation coefficient of  $-0.86$ . Saidi et al. [46] studied the contribution of renewable and nuclear energy in reducing CO<sub>2</sub> emissions in OECD countries. The results showed that a 1 percent increase in renewable energy consumption reduces CO<sub>2</sub> emissions in Sweden by 0.2517%. Nuclear energy was also shown to be impactful in investment in reducing CO<sub>2</sub> emissions in countries such as Canada, Netherlands, Japan, Switzerland, the Czech Republic, and the UK. Imran et al. [47] aimed at examining the effect of the clean energy demand and financing on reducing carbon emissions in 29 economies in Europe and Asia from 2007 to 2020. It was suggested that increasing investment in nuclear energy and green financing can enhance the regional environmental quality. Moreover, the authors found a causal link between fuel imports, nuclear power, and regional growth. In summary, the above-referenced research is in alignment with the findings of this research, which recommends allocating more resources toward investment in hydropower, renewable, and nuclear energy production industries to cut CO<sub>2</sub> emissions and develop a sustainable society.

#### 5. Conclusions

This study focused on the utilization of artificial intelligence and data-driven methods to predict the CO<sub>2</sub> emissions on a national level. Given the limited number of data points available, an enhanced version of multi-task learning (MTL) called weighted multi-task learning (WMTL) was proposed. The WMTL approach aims to extract maximum information from the available data by assigning weights to output tasks and selecting pertinent subtasks, using a devised algorithm. Bayesian optimization is employed to optimize the hyperparameters of the models. The results indicate that the proposed new approach outperforms both the basic and MTL models in terms of accuracy and robustness. The WMTL model obtained an average mean squared error (MSE) of 0.12 across folds, outperforming the MTL model's MSE of 0.15 and the basic model's MSE of 0.21. Additionally, the computational cost of training the new model is significantly lower, at only 20% compared to the other two models. In this study, the interpretation of the selected WMTL model involves a combination of Permutation Feature Importance and Partial Dependency analysis to determine the direction of the relationships between the input features and the output.

The results derived from analyzing the WMTL model suggest its potential as a valuable data-driven tool for creating decision-support systems that can identify effective strategic actions to reduce CO<sub>2</sub> emissions significantly. In this regard, our approach can serve as a fundamental method for forecasting CO<sub>2</sub> emissions in other countries in future research studies. Finally, considering the favorable results obtained from the WMTL model, there is potential for its application in conjunction with the developed algorithm for tabular problems, especially when faced with limited data points.

**Author Contributions:** Conceptualization, M.T. and F.G.; methodology, M.T. and E.G.F.; software, M.T.; validation, M.T., E.G.F., F.G. and M.A.; formal analysis, F.G.; investigation, M.T. and F.G.; resources, F.G.; data curation, F.G.; writing—original draft preparation, M.T. and M.A.; writing—review and editing, F.G. and M.A.; visualization, M.T. and E.G.F.; supervision, F.G.; project administration, F.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** No data were created for this research. Section 2.1 and Table A1 provide information about the historical data used in this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Table A1 provides comprehensive information about the input features and outputs of the models, including their full names and statistical descriptions.

**Table A1.** Description of all input features and outputs of the AI models.

Acronym	Parameter	Min	Max	Average
Input 1: Temperature	Temperature, °C	3.8	6.9	5.8
Input 2: GDPc	GDP Per Capita, USD/person	24,425.0	61,127.0	42,045.1
Input 3: Primary_Energy	Primary energy consumption, TWh	566.1	691.3	624.1
Input 4: C_El	Electricity consumption, TWh	134.4	147.1	142.9
Energy supply (GWh)				
Input 5: S_tot	Total electricity supply	149,718.0	177,982.0	163,721.3
Input 6: S_hydro	Electricity supply from hydro	51,740.0	79,061.0	67,420.5
Input 7: S_pstorage	Electricity supply from pump storage	22.0	565.0	151.0
Input 8: S_nuclear	Electricity supply from nuclear	52,173.0	77,671.0	66,777.2
Input 9: S_CHP	Electricity supply from main activity producer CHP	2290.0	12,721.0	7013.1
Input 10: S_autoCHP	Electricity supply from autoproducer CHP	2650.0	6959.0	4906.2
Input 11: S_wind	Electricity supply from wind	0.0	19,847.0	3999.8
Input 12: S_solar	Electricity supply from solar	0.0	663.0	33.6
Input 13: S_cond_turbines	Electricity supply from condensing turbine	174.0	3869.0	673.7
Input 14: S_gas-turbines	Electricity supply from gas turbines for reserve and others	7.0	147.0	38.4
Input 15: S_import	Electricity supply from import	6102.0	24,286.0	12,709.1
Consumption of fuels in electricity generation (TJ)				
Input 16: CF_o1	No. 1 fuel oil	186.0	4562.0	1105.2
Input 17: CF_o2	No. 2 fuel oil	0.0	1263.0	428.2
Input 18: CF_o23	Nos. 2 and 3 fuel oil	0.0	236.0	14.6
Input 19: CF_o35	Nos. 3–5 fuel oil	0.0	40,433.0	6370.9
Input 20: CF_o4	No. 4 fuel oil	0.0	342.0	64.7
Input 21: CF_o5h	No. 5 and heavier fuel oils	0.0	18,605.0	2526.2
Input 22: CF_coal	Hard coal	983.0	21,802.0	6648.0
Input 23: CF_peat	Peat and peat briquettes	178.0	3834.0	1273.0
Input 24: CF_wood1	Wood briquettes and pellets	0.0	4561.0	1294.5
Input 25: CF_wood2	Wood chips, wood waste, saw dust, etc.	3949.0	25,131.0	13,970.2
Input 26: CF_kerosene	Kerosene	0.0	147.0	32.4
Input 27: CF_diesel oil	Diesel oil	0.0	13.0	5.3
Input 28: CF_NG	Natural gas	1270.0	10,449.0	3408.3
Input 29: CF_biogas	Biogas	0.0	224.0	92.1

Table A1. Cont.

Acronym	Parameter	Min	Max	Average
Input 30: CF_oven_gas	Coke oven gas	186.0	742.0	475.7
Input 31: CF_furnace_gas	Blast furnace gas, incl.	2383.0	8701.0	4708.7
Input 32: CF_liquor	LD gasBlack liquor, spent liquor, tall oil, and pitch oil	0.0	29,895.0	9918.9
Input 33: CF_LPG	Liquid petroleum gas (LPG)	0.0	544.0	147.6
Input 34: CF_nuclear	Nuclear fuel	539,704.0	822,396.0	694,604.4
Input 35: CF_solid_waste	Municipal solid waste	385.0	14,047.0	5334.6
Input 36: CF_other	Other fuels	508.0	5181.0	2360.6
Input 37: CF_fuels	Sum of fuels	612,760.0	894,501.0	755,997.4
Input 38: CF_surplus_steam	Surplus steam	0.0	1185.0	294.2
Input 39: CF_tot_fuels_steam	Sum of fuels and steam	612,760.0	895,351.0	756,296.9
Consumption of fuels for steam and hot water production (TJ)				
Input 40: CFH_o1	No. 1 fuel oil	1471.0	7112.0	3373.2
Input 41: CFH_o2	no. 2 fuel oil	0.0	2996.0	852.4
Input 42: CFH_o23	Nos. 2 and 3 fuel oil	0.0	1986.0	254.3
Input 43: CFH_o35	Nos. 3–5 fuel oil	0.0	22,827.0	5425.0
Input 44: CFH_o4	No. 4 fuel oil	0.0	3161.0	530.6
Input 45: CFH_o5h	No. 5 and heavier fuel oils	0.0	17,585.0	2712.6
Input 46: CFH_coal	Hard coal	2827.0	26,229.0	9739.0
Input 47: CFH_peat	Peat and peat briquettes	3149.0	13,728.0	9153.5
Input 48: CFH_wood1	Wood briquettes and pellets	0.0	22,717.0	13,102.7
Input 49: CFH_wood2	Wood chips, wood waste, saw dust, etc.	13,316.0	77,580.0	48,508.4
Input 50: CFH_kerosene	Kerosene	0.0	83.0	3.2
Input 51: CFH_diesel_oil	diesel oil	0.0	29.0	3.9
Input 52: CFH_NG	natural gas	3707.0	24,036.0	10,266.8
Input 53: CFH_biogas	Biogas	0.0	1626.0	760.0
Input 54: CFH_oven_gas	coke oven gas	115.0	653.0	399.2
Input 55: CFH_furnace_gas	blast furnace gas, incl.	2438.0	3921.0	3032.2
Input 56: CFH_liquor	LD gasBlack liquor, spent liquor, tall oil and pitch oil	0.0	7909.0	3678.8
Input 57: CFH_LPG	Liquid petroleum gas (LPG)	42.0	4636.0	1249.6
Input 58: CFH_solid_waste	Municipal solid waste	14,119.0	56,346.0	29,210.1
Input 59: CFH_other	Other fuels	609.0	21,635.0	8782.3
Input 60: CFH_fuels	Sum of fuels	88,400.0	209,185.0	151,035.2
Emissions of greenhouse gases (kt CO <sub>2</sub> -eqv.)				
Output 1: Em_Total	Total air emissions	23,123.0	44,968.1	34,911.9
Output 2: Em_Agriculture	Emissions from agriculture sector	6714.4	7763.5	7189.6
Output 3: Em_Transport	Emissions from transport sector	16,428.1	21,401.3	19,773.7
Output 4: Em_ElecHeat	Emissions from electricity and district heating	4537.3	11,665.4	6617.7
Output 5: Em_HeatHouse	Emissions from heating of houses and buildings	804.0	9298.1	4458.3
Output 6: Em_Industry	Emissions from industry sector	15,751.9	22,438.5	19,945.8
Output 7: Em_InternationalTransport	Emissions from international transport sector	3725.2	10,191.4	7152.0
Output 8: Em_Offroad	Emissions from off-road vehicles and other machinery	2804.9	3584.2	3324.5
Output 9: Em_Solvent	Emissions from solvent use and other product	489.9	1792.1	1331.0
Output 10: Em_Waste	Emissions from waste	1094.4	3819.8	2661.2

## References

- European Union. *Special Eurobarometer 459 “Climate Change”*; European Union Commission: Brussels, Belgium, 2017; ISBN 978-92-79-70220-4. [\[CrossRef\]](#)
- Capros, P.; Zazias, G.; Evangelopoulou, S.; Kannavou, M.; Fotiou, T.; Siskos, P.; De Vita, A.; Sakellaris, K. Energy-System Modelling of the EU Strategy towards Climate-Neutrality. *Energy Policy* **2019**, *134*, 110960. [\[CrossRef\]](#)
- Salvia, M.; Reckien, D.; Pietrapertosa, F.; Eckersley, P.; Spyridaki, N.-A.; Krook-Riekkola, A.; Olazabal, M.; De Gregorio Hurtado, S.; Simoes, S.G.; Geneletti, D.; et al. Will Climate Mitigation Ambitions Lead to Carbon Neutrality? An Analysis of the Local-Level Plans of 327 Cities in the EU. *Renew. Sustain. Energy Rev.* **2021**, *135*, 110253. [\[CrossRef\]](#)
- National Long-Term Strategies. Available online: [https://Commission.Europa.Eu/Energy-Climate-Change-Environment/Implementation-Eu-Countries/Energy-and-Climate-Governance-and-Reporting/National-Long-Term-Strategies\\_en#national-Long-Term-Strategies](https://Commission.Europa.Eu/Energy-Climate-Change-Environment/Implementation-Eu-Countries/Energy-and-Climate-Governance-and-Reporting/National-Long-Term-Strategies_en#national-Long-Term-Strategies) (accessed on 28 June 2023).

5. Swedish Government. Ett Klimatpolitiskt Ramverk För Sverige (A Climate Policy Framework for Sweden). 2017. Available online: <https://www.government.se/articles/2021/03/swedens-climate-policy-framework/> (accessed on 28 June 2023).
6. Gambhir, A.; George, M.; McJeon, H.; Arnell, N.W.; Bernie, D.; Mittal, S.; Köberle, A.C.; Lowe, J.; Rogelj, J.; Monteith, S. Near-Term Transition and Longer-Term Physical Climate Risks of Greenhouse Gas Emissions Pathways. *Nat. Clim. Chang.* **2022**, *12*, 88–96. [CrossRef]
7. Sognnaes, I.; Gambhir, A.; van de Ven, D.-J.; Nikas, A.; Anger-Kraavi, A.; Bui, H.; Campagnolo, L.; Delpiazzo, E.; Doukas, H.; Giarola, S.; et al. A Multi-Model Analysis of Long-Term Emissions and Warming Implications of Current Mitigation Efforts. *Nat. Clim. Chang.* **2021**, *11*, 1055–1062. [CrossRef]
8. Mansfield, L.A.; Nowack, P.J.; Kasoar, M.; Everitt, R.G.; Collins, W.J.; Voulgarakis, A. Predicting Global Patterns of Long-Term Climate Change from Short-Term Simulations Using Machine Learning. *NPJ Clim. Atmos. Sci.* **2020**, *3*, 44. [CrossRef]
9. Dozić, D.J.; Urošević, B.D.G. Application of Artificial Neural Networks for Testing Long-Term Energy Policy Targets. *Energy* **2019**, *174*, 488–496. [CrossRef]
10. Dey, S.; Reang, N.M.; Majumder, A.; Deb, M.; Das, P.K. A Hybrid ANN-Fuzzy Approach for Optimization of Engine Operating Parameters of a CI Engine Fueled with Diesel-Palm Biodiesel-Ethanol Blend. *Energy* **2020**, *202*, 117813. [CrossRef]
11. Zeng, S.; Su, B.; Zhang, M.; Gao, Y.; Liu, J.; Luo, S.; Tao, Q. Analysis and Forecast of China’s Energy Consumption Structure. *Energy Policy* **2021**, *159*, 112630. [CrossRef]
12. Xiao, X.; Mo, H.; Zhang, Y.; Shan, G. Meta-ANN—A Dynamic Artificial Neural Network Refined by Meta-Learning for Short-Term Load Forecasting. *Energy* **2022**, *246*, 123418. [CrossRef]
13. Ahmadi, A.; Talaei, M.; Sadipour, M.; Amani, A.M.; Jalili, M. Deep Federated Learning-Based Privacy-Preserving Wind Power Forecasting. *IEEE Access* **2023**, *11*, 39521–39530. [CrossRef]
14. Hu, Y.; Man, Y. Energy Consumption and Carbon Emissions Forecasting for Industrial Processes: Status, Challenges and Perspectives. *Renew. Sustain. Energy Rev.* **2023**, *182*, 113405. [CrossRef]
15. Klyuev, R.V.; Morgoev, I.D.; Morgoeva, A.D.; Gavrina, O.A.; Martyushev, N.V.; Efremenkov, E.A.; Mengxu, Q. Methods of Forecasting Electric Energy Consumption: A Literature Review. *Energies* **2022**, *15*, 8919. [CrossRef]
16. Aryai, V.; Goldsworthy, M. Day Ahead Carbon Emission Forecasting of the Regional National Electricity Market Using Machine Learning Methods. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106314. [CrossRef]
17. Kermanshahi, B.; Iwamiya, H. Up to Year 2020 Load Forecasting Using Neural Nets. *Int. J. Electr. Power Energy Syst.* **2002**, *24*, 789–797. [CrossRef]
18. Azadeh, A.; Ghaderi, S.F.; Sohrabkhani, S. A Simulated-Based Neural Network Algorithm for Forecasting Electrical Energy Consumption in Iran. *Energy Policy* **2008**, *36*, 2637–2644. [CrossRef]
19. Sadri, A.; Ardehali, M.M.; Amirnekoeei, K. General Procedure for Long-Term Energy-Environmental Planning for Transportation Sector of Developing Countries with Limited Data Based on LEAP (Long-Range Energy Alternative Planning) and EnergyPLAN. *Energy* **2014**, *77*, 831–843. [CrossRef]
20. Mason, K.; Duggan, J.; Howley, E. Forecasting Energy Demand, Wind Generation and Carbon Dioxide Emissions in Ireland Using Evolutionary Neural Networks. *Energy* **2018**, *155*, 705–720. [CrossRef]
21. Marjanović, V.; Milovančević, M.; Mladenović, I. Prediction of GDP Growth Rate Based on Carbon Dioxide (CO<sub>2</sub>) Emissions. *J. CO<sub>2</sub> Util.* **2016**, *16*, 212–217. [CrossRef]
22. Sun, W.; Xu, Y. Financial Security Evaluation of the Electric Power Industry in China Based on a Back Propagation Neural Network Optimized by Genetic Algorithm. *Energy* **2016**, *101*, 366–379. [CrossRef]
23. Sahraei, M.A.; Çodur, M.K. Prediction of Transportation Energy Demand by Novel Hybrid Meta-Heuristic ANN. *Energy* **2022**, *249*, 123735. [CrossRef]
24. Fürnkranz, J. Decision Tree. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010; pp. 263–267, ISBN 978-0-387-30164-8.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
26. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4171–4186.
27. Shwartz-Ziv, R.; Armon, A. Tabular Data: Deep Learning Is Not All You Need. *Inf. Fusion* **2022**, *81*, 84–90. [CrossRef]
28. Singh, M.; Dubey, R.K. Deep Learning Model Based CO<sub>2</sub> Emissions Prediction Using Vehicle Telematics Sensors Data. *IEEE Trans. Intell. Veh.* **2023**, *8*, 768–777. [CrossRef]
29. Safa, M.; Nuthall, P.L. Predicting CO<sub>2</sub> Emissions from Farm Inputs in Wheat Production Using Artificial Neural Networks and Linear Regression Models “Case Study in Canterbury, New Zealand”. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 9.
30. Khashman, A.; Khashman, Z.; Mammadli, S. Arbitration of Turkish Agricultural Policy Impact on CO<sub>2</sub> Emission Levels Using Neural Networks. *Procedia Comput. Sci.* **2016**, *102*, 583–587. [CrossRef]
31. Yin, L.; Liu, G.; Zhou, J.; Liao, Y.; Ma, X. A Calculation Method for CO<sub>2</sub> Emission in Utility Boilers Based on BP Neural Network and Carbon Balance. *Energy Procedia* **2017**, *105*, 3173–3178. [CrossRef]

32. Ye, H.; Ren, Q.; Hu, X.; Lin, T.; Shi, L.; Zhang, G.; Li, X. Modeling Energy-Related CO<sub>2</sub> Emissions from Office Buildings Using General Regression Neural Network. *Resour. Conserv. Recycl.* **2018**, *129*, 168–174. [[CrossRef](#)]
33. Ma, N.; Shum, W.Y.; Han, T.; Lai, F. Can Machine Learning Be Applied to Carbon Emissions Analysis: An Application to the CO<sub>2</sub> Emissions Analysis Using Gaussian Process Regression. *Front. Energy Res.* **2021**, *9*, 756311. [[CrossRef](#)]
34. Du, J.; Zheng, Q.; Wang, Y. Mid-Term and Long-Term Prediction of Carbon Emissions in Jiangsu Province Based on PCA-STIRPAT Improved GA-BP. In Proceedings of the 2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Shenyang, China, 17–19 November 2021; pp. 1–7.
35. Brito, M.; Pires, C.; Santos, L.; Simonelli, G. Prediction of CO<sub>2</sub> Brazilian Emissions with Scenario Analysis Based on Energy and Environmental Indicators. *Res. Sq.* **2023**. [[CrossRef](#)]
36. Han, Y.; Cao, L.; Geng, Z.; Ping, W.; Zuo, X.; Fan, J.; Wan, J.; Lu, G. Novel Economy and Carbon Emissions Prediction Model of Different Countries or Regions in the World for Energy Optimization Using Improved Residual Neural Network. *Sci. Total Environ.* **2023**, *860*, 160410. [[CrossRef](#)]
37. Zhang, Y.; Yang, Q. An Overview of Multi-Task Learning. *Natl. Sci. Rev.* **2018**, *5*, 30–43. [[CrossRef](#)]
38. Glantz, S.A.; Slinker, B.K. *Primer of Applied Regression and Analysis of Variance*; McGraw-Hill, Health Professions Division: New York, NY, USA, 1990; ISBN 9780070234079.
39. Greenhill, S.; Rana, S.; Gupta, S.; Vellanki, P.; Venkatesh, S. Bayesian Optimization for Adaptive Experimental Design: A Review. *IEEE Access* **2020**, *8*, 13937–13948. [[CrossRef](#)]
40. Wu, J.; Chen, X.-Y.; Zhang, H.; Xiong, L.-D.; Lei, H.; Deng, S.-H. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40. [[CrossRef](#)]
41. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
42. Friedman, J.H.; Popescu, B.E. Predictive Learning via Rule Ensembles. *Ann. Appl. Stat.* **2008**, *2*, 916–954. [[CrossRef](#)]
43. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 972–981.
44. Berga, L. The Role of Hydropower in Climate Change Mitigation and Adaptation: A Review. *Engineering* **2016**, *2*, 313–318. [[CrossRef](#)]
45. Mohsin, M.; Orynassarov, D.; Anser, M.K.; Oskembayev, Y. Does Hydropower Energy Help to Reduce CO<sub>2</sub> Emissions in European Union Countries? Evidence from Quantile Estimation. *Env. Dev.* **2023**, *45*, 100794. [[CrossRef](#)]
46. Saidi, K.; Omri, A. Reducing CO<sub>2</sub> Emissions in OECD Countries: Do Renewable and Nuclear Energy Matter? *Prog. Nucl. Energy* **2020**, *126*, 103425. [[CrossRef](#)]
47. Imran, M.; Zaman, K.; Nassani, A.A.; Dinçã, G.; Khan, H.u.R.; Haffar, M. Does Nuclear Energy Reduce Carbon Emissions despite Using Fuels and Chemicals? Transition to Clean Energy and Finance for Green Solutions. *Geosci. Front.* **2023**, 101608. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.