engineering proceedings

MDPI

Proceeding Paper

A Secure IoT-Cloud Based Remote Health Monitoring for Heart Disease Prediction Using Machine Learning and Deep Learning Techniques †

Sibo Prasad Patro * and Neelamadhab Padhy

Department of Computer Science and Engineering, GIET University, Gunupur 765022, India; dr.neelamadhab@giet.edu

* Correspondence: sibofromgiet@giet.edu; Tel.: +91-94372-34031

† Presented at the 4th International Electronic Conference on Applied Sciences, 27 October–10 November 2023; Available online: https://asec2023.sciforum.net/.

Abstract: The Internet of Things (IoT) refers to a network of interconnected devices as well as technology that enables objects to communicate with one another and the cloud for modern medical treatment. To analyze and handle remotely collected electronic clinical records, it is important to create a disease prediction model with increased accuracy. An RHMIoT framework is proposed in a secure cloud context using lightweight block encryption and decryption approaches. The accuracy levels of cardiac disease are calculated using machine learning and deep learning methods. The ensemble voting classifier provided the greatest accuracy of 95%.

Keywords: IoT; cloud; machine learning; deep learning; severity level

check for updates

Citation: Patro, S.P.; Padhy, N. A Secure IoT-Cloud Based Remote Health Monitoring for Heart Disease Prediction Using Machine Learning and Deep Learning Techniques. *Eng. Proc.* **2023**, *56*, 241. https://doi.org/10.3390/ASEC2023-16580

Academic Editor: Nunzio Cennamo

Published: 4 December 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

Over the last few decades, the cardiovascular disease (CVD) death rate has increased significantly worldwide. In the United States of America, CVD arrest causes one death per minute. According to the World Health Organization (WHO), cardiovascular diseases (CVDs) account for 30% of all fatalities worldwide, with low- and middle-income countries accounting for 75% of these deaths. In India, 25% of people between the ages of 25 and 69 die from CVDs [1]. Cardiovascular disease (CVD) accounted for 42% of all deaths in Australia in 2018, according to the Australian Institute of Health and Welfare (AIHW). Several researchers are working to create a reliable method for the quick detection of heart disease. A proper diagnosis and course of action can save the lives of numerous patients. Heart diseases are diagnosed by a doctor using the patient's medical history, the results of the physical exam, and an analysis of any alarming symptoms. However, this method of diagnosis does not reliably identify patients with heart disease. It is also expensive and computationally difficult to examine.

The patients' health conditions need to be monitored with immense care. Healthcare supports good health and aids in the monitoring of a patient's health state, disease diagnosis, and management, as well as recovery. To achieve a rapid recovery, the health status must be monitored after discharge or after surgery. The machine learning (ML) and deep learning (DL) techniques in healthcare are growing rapidly. Remote healthcare monitoring is a typical healthcare application that helps clinicians in monitoring patients with chronic or acute illnesses in remote places, as well as elderly persons receiving in-home care and hospitalized patients. Due to recent rapid advancements in technology, numerous wearable devices and health monitoring equipment are now readily available on the market. Doctors' jobs are made considerably easier by using contemporary technology like the IoT, ML, and AI. These technologies help to identify the underlying causes of sickness and assess its seriousness using contemporary algorithms.

The IoT refers to physical objects with limited processing and storage capabilities that are connected to the internet. It still has lots of room to grow in the near-future as it faces performance, interoperability, security, and privacy challenges [2]. The IoT has also been established as a benchmark for healthcare systems in which smart objects are used to continuously monitor patients for specific diseases. Smart objects are made up of biomedical sensors that collect health-related data and send them to a physician via the cloud/edge for further analysis. One of the key benefits of the healthcare system is early disease detection. Diabetes and heart disease can be diagnosed earlier using a variety of supervised and unsupervised ML algorithms.

Machine learning and deep learning techniques are proposed in this research to address these problems. The UCI dataset is used for the initial training and validation of machine learning algorithms. Data gathered using sensors in an IoT framework and the prediction of any problems in the health state are calculated in the testing phase. Data from IoT devices are stored in the cloud and are statistically analyzed to determine the accuracy of the prediction percentage.

The goals of this research are as follows:

- A remote-healthcare monitoring system collects the patient's data remotely through medical sensors and performs automated analysis, clustering, and processing, finally visualizing the predicted results to the concerned user.
- The data are transferred to cloud storage using encryption and decryption techniques to prevent unauthorized user access.
- For data preprocessing, the median studentized residual technique and ReliefF are used for the feature extraction process.
- The heart disease severity level and accuracy are calculated on IoT cloud data using ML algorithms.

The remaining part of this paper is arranged as follows. A few similar works are briefly reviewed in Section 2. The experiment's dataset, the ideas included in the suggested strategy, and the suggested framework are all detailed in Section 3. The simulation findings are presented and discussed in Section 4, and the study conclusions are presented in Section 5.

2. Literature Review

This section summarizes several recent related efforts, including a few heart disease prediction models using ML and IoT techniques. The use of machine learning to predict cardiac disease can significantly reduce the number of deaths caused by this heart disease disorder.

Jabeen et al. [3] suggested a Random Forest algorithm for an IoT-enabled healthcare monitoring system. In this study, the Random Forest algorithm was used to predict a variety of diseases, including heart disease, diabetes, and breast cancer, with a maximum accuracy of 97.26% on the dermatology dataset. Kauret al. [4] proposed a system for predicting heart disorders using IoT and machine learning techniques such as J48, NB, Zero R, DT, RF, and BN. IoT sensors were attached to the patient's body to send the patient's medical data into a dataset form. The more accurate results of this research were based on the J48 classifier, which had a 92.56% accuracy rate. Priyadharsan et al. [5] proposed an IoT-based prediction model to track health problems of people using machine learning techniques. The UCI dataset was used for the initial training and validation of machine learning algorithms. The patient's temperature, BP, and heart rate ware measured during the testing phase using an IoT setup. Using the IoT framework, the patient's data were gathered and stored in the cloud for statistical analysis to determine the accuracy of the prediction percentage. Additionally, according to the data, the K-Nearest Neighbor achieved an accuracy of 98.02%. The best performance was obtained when K is 1. Saranya et al. [6] proposed an HD prediction model using IoT using DCNN-FBP algorithms for efficient and effective data analysis. The data were acquired using various wearable gadgets. A deep convolution neural network was used for feature selection. Forward and backward

propagation algorithms were used to exercise the DCNN-FBP algorithm. The proposed model attained 89% accuracy. Ahmed et al. [7] proposed an IoT-based framework using an SVM algorithm. The model experiments evaluated cloud data using the WEKA framework and produced a 97.53% accuracy. The suggested framework was accurately predicted on smaller datasets but for larger datasets, the performance was compromised. Khan et al. [8] proposed an IoT framework using the MDCNN for an accurate heart disease evaluation. In this model, the blood pressure and electrocardiogram data were collected through a smart watch, and the performance of the MDCNN technique was analyzed with deep learning and logistic regression. The model achieved an accuracy of 98.2%.

From the above literature, we found various diverse methods using the IoT and cloud frameworks that were proposed in the literature. Most studies were found with a high validation time, difficulties in partitioning, etc. Few of the researchers were focused on security mechanisms over the cloud to restrict unauthorized access. The literature elucidates the relevant features chosen by various feature selection algorithms to improve the classifier's accuracy.

3. Proposed Methodology

An innovative RHMIoT system architecture is proposed in this research. The framework comprises a sensor layer, transport layer, and application layer. The sensor layer is used to collect the data using various medical IoT sensors. The doctors can monitor a patient from a distance. Using ML-based applications, physicians can continuously analyze their patients' diseases and health status using IoT-medical sensors. The patient's body is fitted with a variety of biological and wearable sensors to collect the patient's vital indicators such as BP, heart rate, and cholesterol levels. For monitoring vital indicators of a patient, various sensors are used in this study, such as a PPG sensor for heart rate monitoring, an optical heart rate sensor for measuring pulse waves, a sensitive stretch sensor for respiration rate, a thermocouple sensor for measuring body temperature, a pressure sensor for monitoring blood pressure, and a pulse oximeter sensor to monitor blood oxygen level.

After the patient data are collected through the sensors, the data are transferred to cloud storage using encryption and decryption techniques to prevent unauthorized users from accessing it. A performance algorithm is used to encrypt sensitive patient data. A lightweight block encryption technique is proposed for confidentiality and security. The raw data are converted to chipper text to prevent unauthorized access. Encryption techniques have a limited impact on IoT security. S-Box is one of the strongest techniques for blocking ciphers. An algorithm is designed for enhancing security with a key-dependent dynamic S-Box and a hyper elliptic curve. In the application layer, heart disease is predicted using ML and deep learning algorithms. The RHMIoT framework is divided into two phases. In phase-1, the patient's risk of HPTN and its severity level are calculated, and in phase-2, heart disease accuracy is calculated using machine learning classifiers.

After retrieving the data from the cloud, the data are processed using various classification techniques. The DM algorithm is applied to the patient's medical data to calculate HPTN and its severity levels. Table 1 shows the classification techniques that are used to categorize the disease of a patient based on their HCL, HPTN level, and HD [9]. The severity level is calculated through various combinations of hypercholesterolemia (HCL), hypertension (HPTN), and severity levels (HPTN1: pre-hypertension; HPTN2: stage I hypertension; HPTN3: stage II hypertension; HPTN4: critical stage hypertension), and heart disease (HD).

In phase-2, the patient's heart disease accuracy is calculated using machine learning and deep learning techniques. After gathering the numerous medical indicators online, the patient's heart condition is monitored. The Framingham dataset is used to train our model which is retrieved from Kaggle. The heart disease dataset contains 4238 records with 16 attributes. The data pre-processing is carried out using feature selection and classification techniques. The patient's heart condition is determined using the training

and testing dataset. After the training process, the IoT medical device sensor data are tested by classifying and contrasting the results. To accurately predict the presence of heart disease, the proposed RHMIoT framework uses the ANN, RF, DT, LR, and VTC. The following section provides a detailed explanation of the training procedure. Figure 1 shows the detailed flow of the proposed RHMIoT framework.

Table 1. Combination of several heart disease disorders for consideration.

Sl. No.	Disease Category
1	No-Disease
2	HCL
3	HPTN-1, HCL
4	HPTN-2, HCL
5	HPTN-3, HCL
6	HPTN-4, HCL
7	Heart Disease, HCL
8	HPTN-1, Heart Disease, HCL
9	HPTN-2, Heart Disease, HCL
10	HPTN-3, Heart Disease, HCL
11	HPTN-4, Heart Disease, HCL

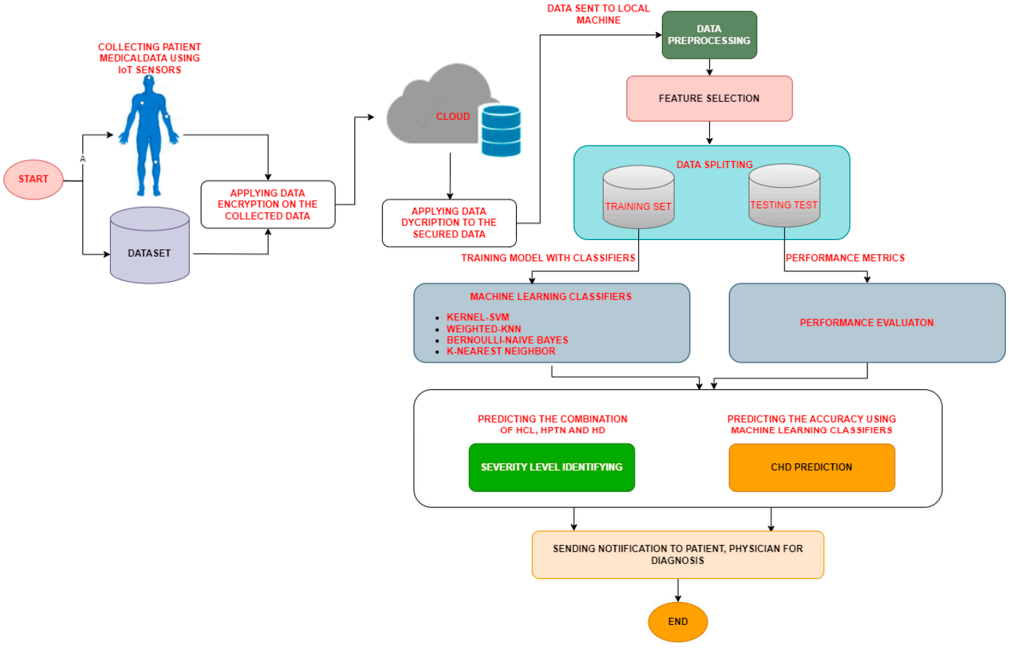


Figure 1. Proposed RHMIoT framework.

3.1. Data Preprocessing

The data preprocessing method is used to replace missing data or remove noise. The missing values of the dataset are identified and updated by the median value of the attribute. A studentized residual technique is used to lower the residuals. It finds the correlation among the features and helps to find the outliers of the given dataset. The preprocessing of data aids in improving heart disease detection. After data pre-processing, the dataset is normalized using a min–max normalization technique using Equation (1).

$$\acute{V}_I = \frac{V_I - MIN_A}{MAX_A_MIN_A}(NEW_MAX_A - NEW_MIN_A)NEW_MIN_A \quad (1)$$

where min (A) returns the min value and max (A) returns the max value of the attribute. After data pre-processing, ReliefF feature selection algorithms are used to identify the appropriate features [10]. It is an effective and general feature estimator and it has the ability to work with multiple classes. Relief assigns a feature score to each feature, which is subsequently utilized to rank and find the highest-scoring features for feature selection. ReliefF can identify features that are most relevant for heart disease prediction in sensor data. The patient's heart condition is determined using the training and testing dataset with an 80:20% ratio.

3.2. Machine Learning Algorithms

The flow of a heart disease prediction model begins with the dataset's input parameters. After data preprocessing, the dataset is passed on to the proposed framework. For comparative analysis, the following techniques are used in this study.

3.2.1. Logistic Regression (LR)

LOR is used fully for binary classification problems. The LOR model defines a predictive variable y in [0, 1]. The LR algorithm is used to analyze the dataset when more than one independent variable performs the result. Using a sigmoid function, LOR transforms the output to achieve the probability value. The sigmoid function calculation is shown in Equation (2):

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \tag{2}$$

The probability estimation of the output of $\sigma(z)$ lies between 0 (normal) and 1 (abnormal). Equation (3) shows the independent features for n in the proposed logistic regression model as $f_1, f_2 \ldots, f_n$:

$$\log \frac{p}{(1-p)} = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \ldots \beta_n f_n \tag{3}$$

3.2.2. Random Forest (RF)

RF is used for constructing different types of DT classifiers. It helps to produce a good result without hyper-tuning. If it is a classification problem, it builds several DTs on different samples and then applies a majority vote to find the best model. We used the proportional random sampling approach to create a fresh training sample set using data from the Framingham dataset. After randomly selecting M feature sets from each training sample, we created B decision trees from the sample set. After dividing the decision tree, the important feature set was selected from the available m features.

3.2.3. Decision Tree (DT)

DT learning is one of the predictive modeling methods that are used in statistics, data mining, and machine learning. Classification trees are tree models with a discrete range of values for the target variable. The branches of these tree structures represent the attributes, and the leaves represent the class labels that are combined to form the class labels. When the target variable can accept continuous values (typically real numbers), regression trees are used instead of decision trees.

3.2.4. Ensemble Voting Classifier

Based on the results of several base models, a voting classifier can forecast the outcome of a vote [11]. The output of each estimator can be voted on separately to build the aggregation criterion. After the three models mentioned in this section are combined, the classification is made using the majority vote of the models (hard voting). For each instance, each model makes a prediction, and the prediction that receives more than half of the votes is the output prediction. Voting classifiers combine the results of various models'

predictions. The class label with the highest number of votes is predicted by adding the predictions for each class label.

The effectiveness and utility of several classification algorithms for the prediction of heart disease are assessed in this study using model construction. The model performance is assessed using confusion matrix parameters.

4. Results and Discussion

The aim of this study is to calculate the severity level of heart disease and the prediction accuracy. In this section, the experiments are conducted in two phases. In phase-1 and phase-2, the severity level and the heart disease accuracy are calculated, respectively. We applied the logistic regression, Random Forest, decision tree, ANN, and ensemble approach to a heart disease test dataset.

The majority voting method uses an ensemble approach to combine various classifiers in order to increase accuracy. LR and RF are recognized as the weak classifiers for the proposed work because they demonstrated extremely low accuracy. When compared to the other algorithms, decision tree performed well and achieved a higher classification accuracy. As seen in Figure 1, the accuracy of weak classifiers was somewhat improved by an ensemble of weak classifiers and strong classifiers using the majority voting technique. Due to the ANN producing very low accuracy, we have not considered it in the following table. Table 2 shows the results of the proposed classifiers.

Table 2. Experimental results.

Classifier	Accuracy	Precision	Recall
Random Forest	90.00%	90.00%	91.00%
Decision Tree	91.00%	92.00%	93.00%
Logistic Regression	89.00%	93.00%	78.00%
Ensemble Voting Classifier	95.00%	93.00%	92.00%

Figure 2 depicts the box plot to show the performance of the proposed classifiers. Using the box plot, we made the decision about which classifiers perform well across multiple measures. The box plot helps us to understand the trade-offs between various performance measures for each classifier and provided support in the choice of the best classifier. Figure 3 depicts the performance metrics and correlation metrics heat map to show the coefficients of correlation between two variables. The variables are each classifier's performance measures such as accuracy, precision, and recall. The correlation coefficient, which runs from −1 to 1, represents the degree and direction of a linear link between two variables.

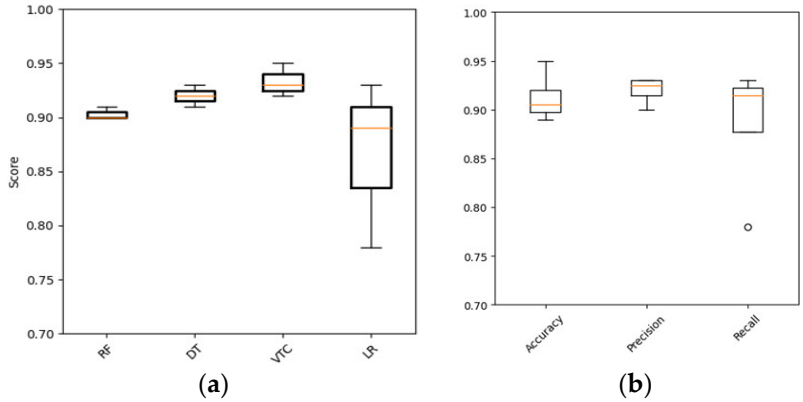
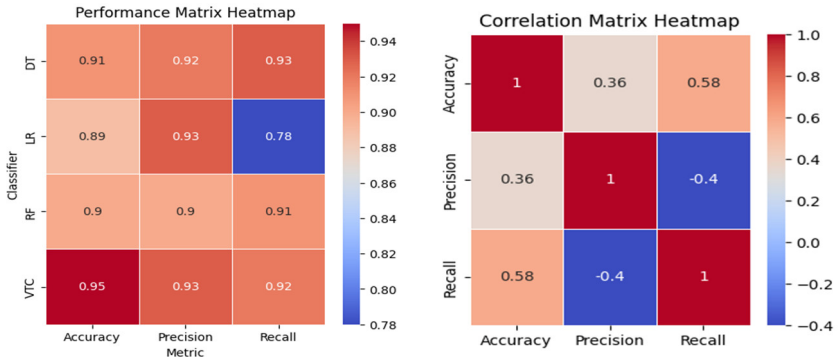


Figure 2. (**a**) Classifier performance metrics. (**b**) Performance evaluation metrics.



(**a**) (**b**)

Figure 3. (**a**) Performance matrix heatmap. (**b**) Correlation matrix heatmap.

5. Conclusions

Due to the rapid increase in CVDs, remote health monitoring has become more popular over the globe. This technique helps patients with diagnoses at home or in a remote area. In this chapter, a lightweight block encryption and decryption technique is provided for a secure RHMIoT. A variety of medical IoT sensors are used to gather data to test the suggested model in two phases. Phase-1involves calculating the severity level of HCL, HPTN, and HD using data mining algorithms. Phase-2 involves calculating the accuracy of heart disease. The outcomes were determined using several performance matrices. In comparison to other ML and DL techniques, the performance of the ensemble voting classifier model provided the greatest accuracy of 95.00%.

We will focus on how to obtain a dataset that is ready for the upcoming machine learning model in our upcoming work. We will try to improve the speed and precision of our model by making a few dynamic adjustments in accordance with the requirements of the user.

Author Contributions: This study was designed, directed, and coordinated by N.P. He also verified the analytical methods, conceived the original idea, and supervised the findings of this work. S.P.P. developed the model, preparation of images, theoretical formalism, data collection, performed the analytic calculations, computational framework, analyzed the data, carried out the experiment, interpretation of the results, analysis of the results, visualization, implementation of the research, and took the lead in writing, reviewing and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available in "UCI" at https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset (accessed on 5 February 2023).

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

References

1. Roth, G.A.; Mensah, G.A.; Johnson, C.O.; Addolorato, G.; Ammirati, E.; Baddour, L.M.; GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **2020**, *76*, 2982–3021. [CrossRef] [PubMed]
2. Kumar, S.; Tiwari, P.; Zymbler, M. Internet of Things is a revolutionary approach for future technology enhancement: A review. *J. Big Data* **2019**, *6*, 1–21. [CrossRef]
3. Jabeen, F.; Maqsood, M.; Ghazanfar, M.A.; Aadil, F.; Khan, S.; Khan, M.F.; Mehmood, I. An IoT based efficient hybrid recommender system for cardiovascular disease. *Peer Netw. Appl.* **2019**, *12*, 1263–1276. [CrossRef]
4. Kaur, P.; Kumar, R.; Kumar, M. A healthcare monitoring system using random forest and internet of things (IoT). *Multimed. Tools Appl.* **2019**, *78*, 19905–19916. [CrossRef]

5. Priyadharsan, D.J.; Sanjay, K.K.; Kathiresan, S.; Karthik, K.K.; Prasath, K.S. Patient health monitoring using IoT with machine learning. *Int. Res. J. Eng. Technol. (IRJET)* **2019**, *6*, 876.
6. Saranya, M.; Preethi, R.; Rupasri, M.; Veena, S. A survey on health monitoring system by using IOT. *Int. J. Res. Appl. Sci. Eng. Technol.* **2018**, *6*, 778–782. [CrossRef]
7. Ahmed, F. An Internet of Things (IoT) application for predicting the quantity of future heart attack patients. *Int. J. Comput. Appl.* **2017**, *164*, 36–40. [CrossRef]
8. Khan, M.A.; Abbas, S.; Atta, A.; Ditta, A.; Alquhayz, H.; Khan, M.F.; Naqvi, R.A. Intelligent cloud based heart disease prediction system empowered with supervised machine learning. *Comput. Mater. Contin.* **2020**, *65*, 139–151.
9. Patro, S.P.; Padhy, N. An RHMIoT Framework for Cardiovascular Disease Prediction and Severity Level Using Machine Learning and Deep Learning Algorithms. *Int. J. Ambient Comput. Intell. (IJACI)* **2022**, *13*, 1–37. [CrossRef]
10. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [CrossRef] [PubMed]
11. Patro, S.P.; Padhy, N.; Sah, R.D. An improved ensemble learning approach for the prediction of cardiovascular disease using majority voting prediction. *Int. J. Model. Identif. Control.* **2022**, *41*, 68–86. [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.