ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

ULTRAVIOLET INDEX ANALYSIS AND FORECASTING USING DEEPLEARNING METHODOLOGIES FOR BENGALURU CITY

Sri Vishnu D[1], Tarun Srivatsa V S[2], Merin Meleet[3]
[1]*Information Science and Engineering, R.V. College of Engineering, Bengaluru, India*
[2]*Information Science and Engineering, R.V. College of Engineering, Bengaluru, India*
[3]*Information Science and Engineering,R.V. College of Engineering, Bengaluru, India*

ABSTRACT
The ultraviolet index is an international standard metric for measuring the strength of the ultraviolet radiation reaching Earth's surface at a particular time, at a particular place. Major health problems may arise from an overexposure to such radiation, including skin cancer or premature aging, just to namea few. Hence, the goal of this work is to make use of Deep Learningmodels to forecast the ultraviolet index at a certain area for futuretimesteps. With the problem framed as a time series one, candidate models are based on Recurrent Neural Networks, a particular class of Artificial Neural Networks that have beenshown to produce promising results when handling time series. Inparticular, candidate models implement Gated Recurrent Unit (GRU) Memory networks, with the models' input ranging from uni to multi-variate. The used dataset was collected from Open Weather Map API. On the other hand, the models' output followsthe approach to forecast UV index for future timestep. The obtained results strengthen the use of the Gated Recurrent Unit (GRU) network to handle time series problems, with the best candidate model achieving high performance and accuracy for ultraviolet index forecasting.

KEYWORDS— ***Gated Recurrent Unit (GRU), Long short term model (LSTM), Artificial Neural Networks, Recurrent NeuralNetworks.***

I. INTRODUCTION

Over the years, the increase of UV radiation reaching Earth's surface has been associated with increased rates of skin cancers, particularly melanomas . Indeed, information regarding UV index variations can be essential for the human being. Despite being harmful in high concentrations, Knowing, beforehand, when the UV index will achieve high or extreme values is of the utmost importance as it allows one to adjust his behaviour and avoid risky moves. Ultraviolet (UV) index is a standard metric used to express the magnitude of UV radiation reachingEarth's surface at a particular time, at a given region. Ozone in the stratosphere, also known as "good" ozone, protects life from harmful UV radiation. However, due to the burning of

fossil fuels, which releases carbon into the atmosphere, the ozone layer has become thinner, leading to dangerous UV radiation reaching Earth's surface. UV radiation is also essential for humanity.

The World Health Organisation and the World Meteorological Organisation proposed a standardised global UV Index scale for a better perception of which concentrations lead to harmful UV radiation. UV index values between 0 and 2 are of low risk; between 3 and 5 are of moderate risk; between 6 and 7 start carrying some risk; between 8 and 11 are very dangerous to the human being; and values higher than 11 are of extreme danger. The goal of this work is to make use of Deep Learning models to forecast the UV index at a certain area for several future timesteps, in particular for the next hour. With the use of Deep Learning models, it becomes possible to forecast future time points in a given scope. Being this a time series problem, uni and multivariate Gated Recurrent Unit (GRU), a subset of Recurrent Neural Networks (RNNs), were conceived and evaluated, with the goal being to forecast the UV index. The objectives of this project include**:**1. Prediction of the UV index of next hour with highest viable accuracy. 2 Considering the temperature as a factor which affects the UV index for prediction. 3 Analysing the effect of UV index on different time stamps.

The higher the UV Index score , the greater the amount of potential skin and eye damaging radiation. At high UV Index levels (greater than 6), significant damage can occur in just a few minutes . Sunburn is the result of overexposure to the UV rays from the sun. Despite harmful in high concentrations, Knowing, beforehand, when the UV index will achieve high orextreme values is of the utmost importance as it allows one to adjust his behaviour and avoid risky moves.

II. LITERATURE SURVEY

A wide range of relevant research papers across platforms were discovered and analysed upon their techniques and different models primarily for object detection and classification.

1. MULTI-STEP ULTRAVIOLET INDEX FORECASTING USING LONG SHORT-TERM MEMORY NETWORKS -

Objectives of the work carried out in this paper was to implement Long Short-Term Memory networks, with the models' input ranging from unito multi-variate to forecast the UV index. And results obtained was the model with the best accuracy in the prediction of the UV index was the Recursive Multi-Step Multi-Variate model with a RMSE of 0.306 and a MAE of0.249, which depict that it is possible to forecast. Were able to forecast for three coming days, recursively [1].

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

2. PATTERNS OF ULTRAVIOLET RADIATION EXPOSURE AND SKIN CANCER RISK: THE E3N-SUNEXP STUDY –
This paper was aimed to quantify the associations between various UV exposures and the risks of melanoma, BCC, and SCC and to examine the patterns thatare the most predictive of skin cancer risk . They concluded that Positive linear relationships between a history of sunburns before 15 years and at 15–25 years and the risks of melanoma Found no statistically significant association between ever use of an indoor tanning device and skin cancer risk [2].

3. DEEP LEARNING FOR TIME SERIES CLASSIFICATION: A REVIEW –
The Objective of the work carried out in this paper was to perform an exhaustive study of DNNs for TSC. Paper concludes that ResNet is the best architecture with FCN following as second best and provides an open source deep learning framework to the TSC community where authors implemented each of the compared approaches and evaluated them [3].

4. ACCURATE SURFACE ULTRAVIOLET RADIATION FORECASTING FOR CLINICAL APPLICATIONS WITH DEEP NEURAL NETWORK –
This paper aims to develop a deep learning model for UV radiation prediction which achieves minimal (around 10%) error for 24-h forecast and 13–16% error for 7-day up to 4-week forecast. And results obtained was initially adapted an encoder–decoder architecture, which can effectively capture relationship between sequence data, to develop an artificial neural network model for forecasting next-day surface UV radiation and also due to overfitting problem of DNNs, In the context of UV forecasting, this dictates that the model must be retrained with data from particular weather station in order to be usable for that geographic region [4].

5. OPTIMAL SUNSCREEN USE, DURING A SUN HOLIDAY WITH A VERY HIGH ULTRAVIOLET INDEX, ALLOWS VITAMIN D SYNTHESIS WITHOUT SUNBURN –
The Objective of the work carried out in this paper was to assess the ability of two intervention sunscreens to inhibit vitamin D synthesis during a week- long sun holiday. The conclusion obtained from this paper was Sunscreens may be used to prevent sunburn yet allow vitamin D synthesis. A high UVA- PF sunscreen enables significantly higher vitamin D synthesis than a low UVA- PF sunscreen because the former, by default, transmits more UVB than the latter [5].

6. IMPROVEMENT OF THE 24 HR FORECAST OF SURFACE UV RADIATION USING AN ENSEMBLE APPROACH –
Paper focuses on the performance of the 24 hr forecast for both the UVI and the maximum duration of safe exposure (DSE) intraday variability. New algorithm is proposed to minimize the forecast error by using the ensemble approach and the offline bootstrap resampling ofcoefficients describing the cloud attenuation of solar UV. which results in Using the random CMF coefficients by the offline bootstrap resampling of the initial CMF matrix improves further the model/observation agreement for the whole period of the UV forecast. Forecasts are better than those obtained by a fixed (for the whole period of analysis) member of the UV forecast ensemble [6].

7. TIME SERIES FORECASTING USING LSTM NETWORKS: A SYMBOLIC APPROACH -
In This paper implementation was done by a combination of a recurrent neural network with a dimension-reducing symbolic representation is proposed and applied for the purpose of time series forecasting.This will significantly speed up the training phase and reduce the model's sensitivity to the hyper parameters and initial weights. Which got results as The time series is of sufficient length, the combined approach can lead to significant speed up of the training phase without degrading the forecast accuracy,whilst reducing the sensitivity to certain hyper parameters
[7].

8.AIR POLLUTANT CONCENTRATION PREDICTION BASED ON GRU METHOD - Paper aims To construct a GRU (gated recurrent units)model to predict the concentration of air pollutants according to the characteristics of high-dimensional and complex air pollution time series.The author experiments using PM2.5concentration information per hour in Beijing were conducted to verify the effectiveness of the GRU model in the field of air quality prediction (NO2). The new predictionmethod presented in this paper is of great significance for the statistical prediction of air quality [8].

9.DEEP BiLSTM-GRU MODEL FOR MONTHLY RAINFALL PREDICTION: A CASE STUDY OF SIMTOKHA, BHUTAN –
This paper talks about pattern recognition to predict precipitation. Bidirectional Long Short Term Memory (BiLSTM) and Gated Recurrent Unit (GRU)-based approach for monthly prediction . The proposed model performed uniformly better than vanilla versions of all the deep learning techniques under study. The MSE score of 0.01 achieved by our model was 41.1% better compared to the next best score of 0.13 provided by LSTM [9].

III. METHODOLOGY

Forecasting is a type of prediction activity which requires mathematical functions to learn from the dataset and then be able to do the required action. Deep neural networks were usedto learn from the formatted dataset.

A deep neural network (DNN) is an ANN with multiple hidden layers between the input and output layers. Similar to shallow ANNs, DNNs can model complex non-linear relationships. Hence they were particularly chosen to be used as the model for forecasting Ultraviolet Index.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

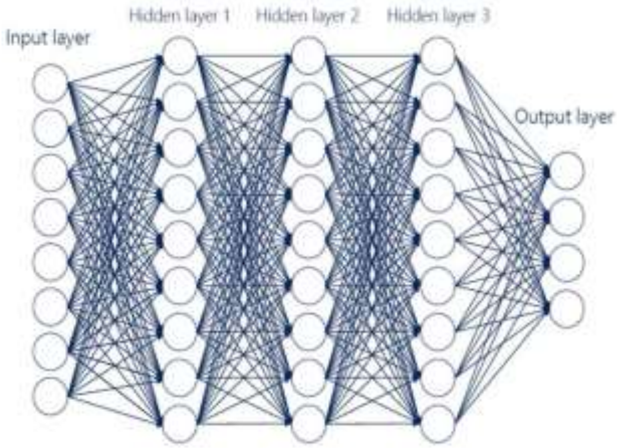


Fig. 3.1: A typical deep neural network showing an input layer, a set of hidden layers along with the output layer. [9]

A typical neural network may recognise complex non linear patterns, yet they are not very efficient at capturing patterns in time series data. Memory based functionality is required for the model to recall from various previous inputs. RNNs constitutea class of artificial neural networks where the evolution of the state depends on the current input as well as the input atprevious timesteps. This property makes it possible to carry outcontext-dependent processing, allowing long-term dependencies to be learned. A recurrent network can have connections that return from the outgoing nodes to the incoming nodes, or even arbitrary connections between any nodes.

But, recurrent neural networks have the problem of vanishing gradient which has been identified by previously conducted research works. Hence the RNNs are modified to form a Longshort-term memory (LSTM) network. LSTMs are a type of RNN architecture. Unlike a traditional neural network, this architecture is used to learn from experience how to classify, process and predict time series, as is the case with this study

[10].

LSTMs contain information outside the normal flow of the recurring network, more specifically in a gated cell. Information can be stored, written or read from a given cell, in an approach similar to data in a computer's memory. These networks are used to process, predict and classify based on time series [10].

Dataset:-Hourly weather data of Bengaluru city. Timestamp starts from 2009-01-01 upto 2020-01-01. Total of 96432 entries were present with no rows containing any null value.

The dataset was first studied for the patterns that can be exhibited based on the timestamp. Several of them were found.

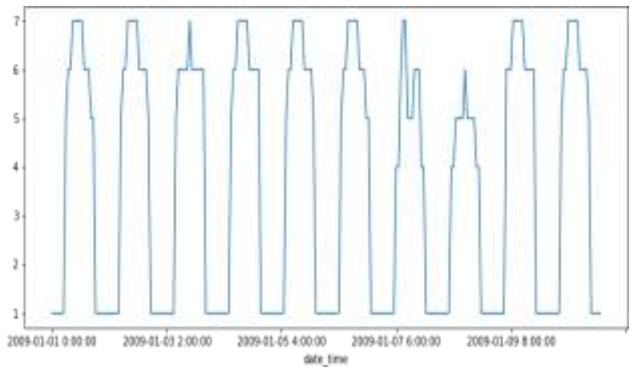


Fig. 3.2: Hourly variation of UV index. It can be observed that there is a recurrent pattern between days

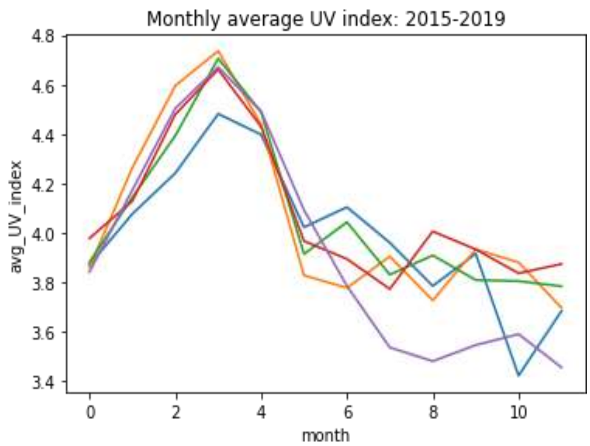


Fig. 3.3: Monthly average UV index over several years. Variance is lower in summer months than that of winter months. UVI peaks in April. Similar pattern is followed everyyear

Several experiments were carried out to find the best combination of hyperparameters for both the uni-variate and the multi-variate approaches. Performance was compared in terms of error-based accuracy, for both the uni-variate and multi-variate candidate GRU models. Regarding the first, it uses only one feature as input, unlike the second which takes into account several features. In fact, the uni-variate models use only the UV index value feature. On the other hand, the multi-variate model uses the UV index value as well as the month and day features, giving a stronger temporal context to the network.

Following are the hyperparameters of the model:
Epoch count: 5, Optimizer: ADAM, Learning Rate: between 0.0001 and 0.001, Evaluation Metric: Root Mean Squared Error (RMSE)

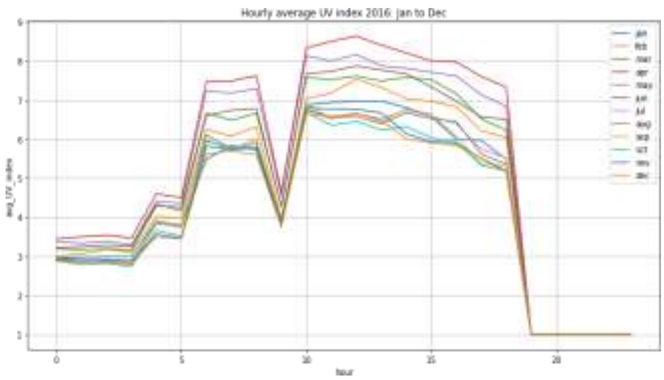


Fig. 3.4: hourly average variation of UVI in different monthsin 2016

The testing part of the train-test split data was used to test the accuracy of each of the models. The input for any model must be given in the same form as that of the input layer of the neural network. There will be no adjustment of weights during the testing process. The model performs the computation in order to give out a predicted value. Then the obtained values are mapped and plotted together for convenient comparison.

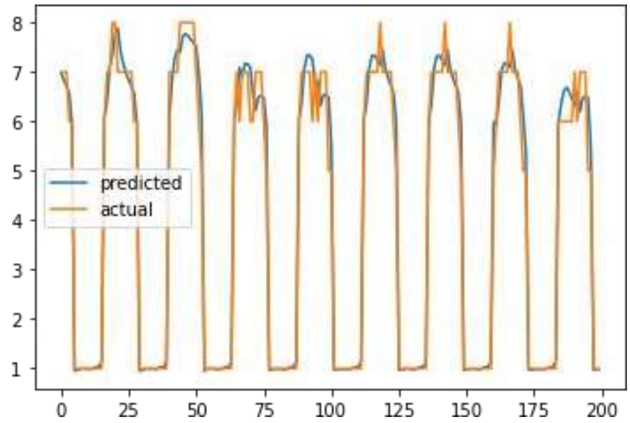


Fig. 4.1: A sample of test data that is plotted along with the values predicted by the second model, ie, the Timestamp model

Test Cases

1. Univariate model was constructed as the most basic model. This model is fast but uses no features or timestamp and hence results in higher Error value over the epochs of training.
2. Timestamp is the most important factor additional to the UVI value because UV Index tends to follow the daily andmonthly patterns. Hence addition of hour of the day and day ofthe year values drastically reduced the training error, hence improving the model.
3. Since temperature values showed maximum correlation with UVI values, it was included as a feature in order to experiment with the model. It tended to slightly improve the accuracy, proving its significance, yet it did not turn out to be as effective as the addition of timestamp was.

RESULTS AND ANALYSIS

Being this a time series forecasting problem, one particular time series cross validation was used, being entitled as TimeSeriesSplit. For each prediction of each split of this cross validator, RMSE and MAE were calculated to be able to evaluate the best set of parameters. The experiments carried out for both uni and multi-variate candidate models made it clear that a stronger temporal context results in an overall decrease of both error metrics eventhough the best uni-variate models have a lower MAE than thebest multi-variate one.

Since both metrics are in the same unit of measurement as the UV index, an error of 0.3 shows that it is possible to forecast, very closely, the expected UV index for the next three days. In the multi-variate model, the inclusion of the month and day yields more accurate predictions in comparison to the uni- variate one. Interestingly, the number of inputs of the model isdirectly proportional to the number of timesteps, i.e., more features as input lead to an increase of the number of timestepsthat are required to build a sequence.

Fig. 4.2: Epochs vs loss and RMSE for UNIVARIATE MODEL. RMSE: 1.3098

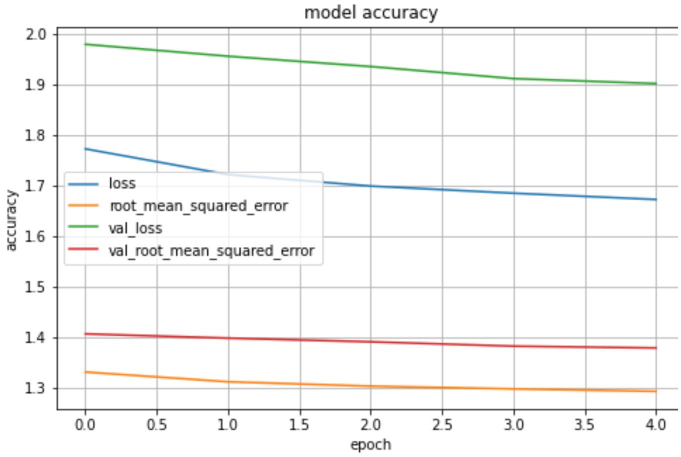
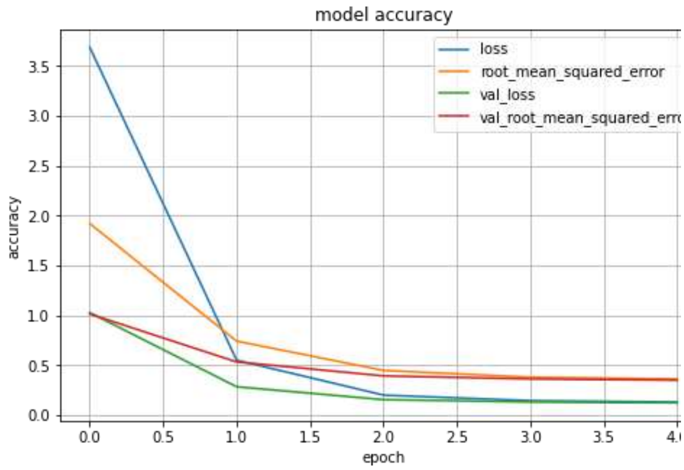


Fig. 4.3: Epochs vs loss and RMSE for TIMESTAMP MODEL. RMSE: 0.3606



ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 8| Issue: 7| July 2022|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2022: 8.205 || ISI Value: 1.188

Fig. 4.4: Epochs vs loss and RMSE for TIMESTAMP MODEL with TEMPERATURE. RMSE: 0.3213

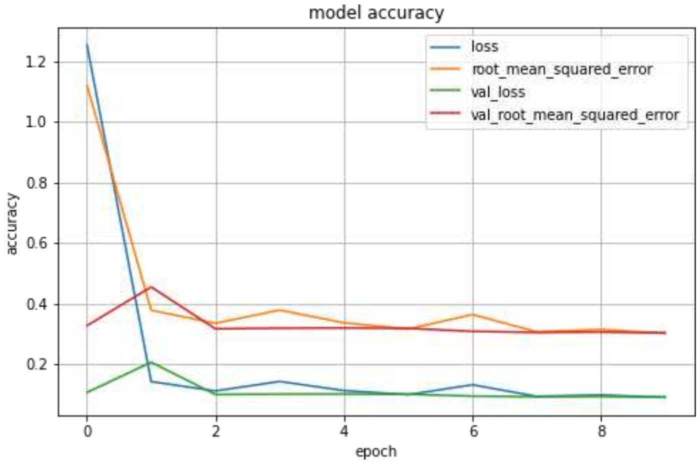
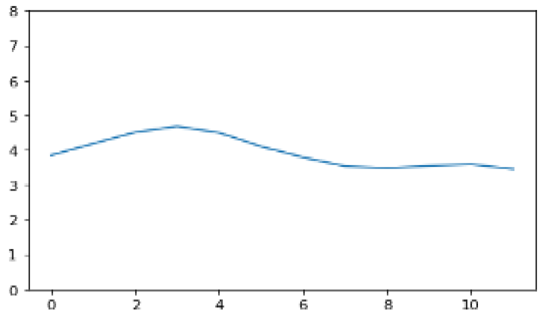


Fig. 3.10: Annual average of UVI. A near flatline observed.



Conclusion

Over the past few years, skin cancer prevention campaigns have increased worldwide. Knowing that exposure to ultraviolet radiation is one of the main causes for such disease, forecasting the UV index assumes particular importance. Hence, this study focused on using deep learning models, in particular GRU based neural networks, to forecast the UV index for the next hour. Multiple experiments were performed, using a wide combination of hyperparameters for all the candidate models. The model with the best accuracy in the prediction of the UV index was the Recursive Multi-Step Multi-Variate model.

INNOVATION

The innovative step in the project was to make use of temperature variation pattern along with Ultraviolet index pattern in order to predict the UV index of the following hours and also to obtain positive results for doing the same. It is so because there exists statistical evidence from the same dataset, i.e, high positive linear correlation of 0.73 between the above two weather factors.

LIMITATIONS

The model's limitations can be evaluated with respect to the algorithm used and also with respect to the dataset. Time seriesdata may have a better algorithm which can be explored. The improvement was flat after 5 epochs. There is also a limitationwith the way intensity of ultraviolet rays is transformed into UV index which is an integer between 1 and 10 in the dataset. Hence, a model that works with classification algorithms mustalso be experimented with.

FUTURE SCOPE

Future work will consider the inclusion of more input features such as the humidity, ozone levels and the position of the sun expressed in terms of solar zenith angle. This addition is expected to further improve the accuracy of forecasting UV index on daily or hourly basis. Further reduction in the RMSE value will help in optimum forecasting and will also allow a higher rate of incorporating machine learning and deep learning techniques for various forecasting problems.

REFERENCES

1. *Pedro Oliveira, Bruno Fernandes, Cesar Analide, Paulo Novais,"Multi-step Ultraviolet Index Forecasting using Long Short-Term Memory Networks", 2021, The book of Distributed Computing and Artificial Intelligence, 17th International Conference.*
2. *Savoye I, Olsen CM, Whiteman DC, Bijon A, Wald L, Dartois L, Clavel-Chapelon F, Boutron-Ruault MC, Kvaskoff*
3. *M. "Patterns of Ultraviolet Radiation Exposure and Skin Cancer Risk" :the E3N-SunExp Study. J Epidemiol. 2018 Jan 5;28(1):27-33. doi: 10.2188/jea.JE20160166. Epub 2017 Nov 25. PMID: 29176271; PMCID: PMC5742376.*
4. *Sunscreen use, during a sun holiday with a very high ultraviolet index, allows vitamin D synthesis without sunburn. Br J Dermatol. 2019 Nov;181(5):1052-1062. doi: 10.1111/bjd.17888. Epub 2019 May 24. PMID: 31069787; PMCID: PMC6899952.*
5. *Jakub Guzikowski, Aleksander Pietruczuk, Piotr S. Sobolewski , "Improvement of the 24 hr forecast of surface UV radiation using an ensemble approach " 2019 , Republic of Poland; National Science Centre.*
6. *Steven Elsworth, Stefan Guttel, "Time Series Forecasting Using LSTM Networks: A Symbolic Approach", 2020, arxiv.org*
7. *Xinxing Zhou et al 2019 J. Phys.: Conf. Ser. 1168 032058*
8. *Chhetri, M.; Kumar, S.; Pratim Roy, P.; Kim, B.-G. Deep BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan. Remote Sens. 2020, 12, 3174.*
9. *Structure of Neural Network: Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation vol.9(8), 1735–1780 (1997)*
10. *Ismail Fawaz, H., Forestier, G., Weber, J. et al. Deep learning for time series classification: a review. Data Min Knowl Disc 33, 917–963 (2019).*
11. *[Raksasat, R., Sri-iesaranusorn, P., Pemcharoen, J. et al. Accurate surface ultraviolet radiation forecasting for clinical applications with deep neural networks. Sci Rep 11, 5031 (2021).*
12. *Young AR, Narbutt J, Harrison GI, Lawrence KP, Bell M, O'Connor C, Olsen P, Grys K, Baczynska KA, Rogowski-Tylman M, Wulf HC, Lesiak A, Philipsen PA. Optimal*