How to Cite:
Minu, M. S., Aaman, S. O., Reddy, C. D., & Charan, V. G. (2022). A generalized and finely tunable rainfall prediction based on optimized forecasting model. *International Journal of Health Sciences*, *6*(S1), 7613–7642. https://doi.org/10.53730/ijhs.v6nS1.6645

A generalized and finely tunable rainfall prediction based on optimized forecasting model

M. S. Minu
Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology. Ramapuram Campus, Chennai, India
Corresponding author email: msminu1990@gmail.com

Sheikh Owais Aaman
Department of Computer Science and Engineering, SRM Institute of Science and Technology. Ramapuram Campus, Chennai, India

C. Dinesh Reddy
Department of Computer Science and Engineering, SRM Institute of Science and Technology. Ramapuram Campus, Chennai, India

Vukka Gurusai Charan
Department of Computer Science and Engineering, SRM Institute of Science and Technology. Ramapuram Campus, Chennai, India

***Abstract*---**Agriculture is the most important factor in India for survival. Rainfall is the most critical factor for agriculture. A new approach for designing the network structure in a deep neural network (DNN) based forecasting model for rainfall prediction is presented. The number of atmospheric parameters determines whether the rain will fall. We suggest a systematic approach for analysing various parameters affecting rainfall in the atmosphere in addition to identifying seasonal and diurnal variables, the ground-based weather features Temperature, Relative Humidity, Dew Point, Solar Radiation, Precipitable Water Vapor (PWV) are also identified, and a detailed feature correlation study is presented. Forecasting rainfall in non-monsoon sessions required input values of temperature, wind velocity, humidity, and cloud cover. Among the parameters used to evaluate results were MSE (mean square error), CCI (correlation coefficient), COE (coefficient of efficiency), and MAE (mean absolute error). By using the proposed approach, the proposed scheme establishes that deep learning (DL) can be effective in improving rainfall prediction quality, exploiting the potential of neural network (NN) in using a weather simulator. In terms of daily

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 18 Feb 2022, Manuscript revised: 09 March 2022, Accepted for publication: 27 April 2022

precipitation characteristics, this various model is applied to 12 different sites under various climatic reinforcements. Currently, the forecasting of weather is based on computer-generated models. To forecast better form Deep Neural Networks (DNN) are widely used to predict the future.

***Keywords*---**DNN (Deep Neural Networks), DP (Deep Learning), atmospheric parameters, Precipitable Water Vapor.

Introduction

Rainfall is necessary for agriculture to succeed. It also enables water resources to grow. In the past, rainfall information helped farmers better manage their crops, leading to economic growth in the country.Hydrology and meteorology are highly dependent on the rainfall ecosystem. Climate factors that influence precipitation include air temperature, sunlight, air humidity, wind speed, and precipitation frequency. Precipitation is the consequence of multi-scale air framework association, and it is impacted by numerous natural elements, for example, nuclear energy, stream field, and territory [1] the forecasting process should be an integral part of decision-making. Based on the most current forecasting techniques and objective evidence, we make weather forecasts. Weather forecasts include numerical models and these factors:A historical view of the weather, Current weather conditions, A historical look at the climatic conditions. There are three types ofatmosphericmodels [2][3].Environmental models are further sub-isolated into airflow models (atmospheric circulation), mathematical models, and climatic models [4].To estimate precipitation on a given day, we use the environmental course model based on barometric data.. Environment models then again conjecture precipitation utilizing climatic elements. Presently, numerical models are used to make predictions for medium-sized rainfalls. But most atmospheric models are computationally intensive. The task is quite challenging. The problem has been tackled through the use of machine learning and artificial intelligence. [5].

Related Work

Numerous studies have been conducted on rainfall prediction. We discuss a few of the approaches we consider here. Rainfall prediction is quite challenging due to dynamic changes in atmospheric processes. Researchers from several fields struggle to create prediction models for rainfall that is accurate, including weather data mining, environmental machine learning, functional hydrology, and numerical forecasting [6]. The problem was solved using a combination of AI, ML, and Deep Learning. In Machine Learning methods, such as Backpropagation Neural Networks with feed-forward neural networks, layer recurrent and cascade networks have been studied [7]. We investigated three ANNs from three different real-world outcomes Several types of neural networks are multilayer feed-forward networks, time-delay neural networks, and partial recurrent neural networks [8]. The ensemble decomposition principle is used to model rainfall forecasts daily. Ensemble Empirical Mode Decomposition (EEMD) is used to break up the original data into its simplest components. The Feed-Forward Neural Network

(FNN) is also employed in the forecasting process. As of today, most weather forecasting methods work by using generative approaches, such as numerical simulations of weather systems or time-series analysis such as ARIMA models, and Support vector machines or Artificial Neural Networks for simple classifiers [10]. Forecasting rainfall most commonly uses feed-forward neural networks that are trained with backpropagation and that employ multilayer perceptron's (MLPs) [14]. Researchers have shown that artificial neural networks are effective in the detection of both high and low rainfall events [15]. In an attempt to overcome the curse of dimensionality, statistical models rely on strong assumptions, like spatial independence. Machine learning, on the other hand, applies deep learning to predict the future. Therefore, if a model performs well with the training data, then it will also perform well when compared with the testing data.

Existing System

As the amount of historical data collected grows, so does the demand for business intelligence applications that include automatic time series forecasting. There is no universal time series modelling approach [9]. A combination of several forecasting methods may be viewed as a compromise. The meta-learning approach, using two separate random forest regression models, ranks 22 univariate forecasting methods and recommends ensemble size based on 390 time-series features. An ensemble of the best forecasting methods is subsequently constructed, and forecasts are pooled either using a simple or weighted average [11]. Overall, the best results were obtained using weighted pooling with an asymmetric mean absolute percentage error of 9.21% versus 11.05% when the Theta method was used.

Drawbacks of Existing System:

- Can't learn the relation between factors.
- Factors such as features should not be taken into consideration
- Maximizes the complexity of the problem
- Prone to Errors
- As the number and amount of data grow, these designs have a problem with scalability

Proposed System

As long as the parameters or attributes are constant, Deep Learning models can provide predictions based on past data, regardless of whether it's district data or state data. States are also averaged based on data from separate districts. So, if a model performs well with data then it is likely to show the same for the entire state. The variables are scaled from 0 to 1 to avoid scaling problems. This is done via the MinMax Scaler class in the sklearn preprocessing library. Le is a part of the dataset after scaling across-validation tests are performed on both the training and testing datasets to assess the predictive capabilities We consider seven-folds of approximately equal size, and the model has been trained repeatedly on six-folds (k1) after which the standard performance is determined.

In addition, the neural network receives one-dimensional vector forms of input data, followed by a flattened layer. A layer of dense, fully connected layers is then added with an activation function. A layer of input neurons is usually larger than an output layer. To accomplish this, neural networks are allowed to think broader before convergence. Setting the architecture of the model requires specifying several parameters. Predictions are produced in the output layer by fully connected layers.

Advantages of Proposed System

- Outstanding Learning Capabilities
- Effectively improve prediction accuracy.
- Reduces the consumption of hardware resources
- Lowering the Complexity Threshold
- Low computational cost

Existing Methodology

Multi-Layer Perception MLP:

An artificial neural network's most useful component is the multilayer perceptron. It was initially called "perceptron" due to its resemblance to a single neuron in an early neural network model. Biological brain models are used to solve computation-intensive tasks like the ones in machine learning, such as predicting outcomes in the future [12]. As opposed to creating realistic models of the brain, the aim is to create robust algorithms and data structures that can be used for modeling difficult problems.

Neural networks are powerful because they can learn the pattern in your training data and then how to relate it to the output variable you want to predict. Essentially, neural networks learn how to map data [13]. Their mathematical capabilities allow them to learn any mapping function and to operate as a universal approximation algorithm. Because they are hierarchical and layered, neural networks can predict. As it learns characteristics at various scales and resolutions, the data structure can combine them into higher-order features. A line can combine with a collection of lines to become a shape.

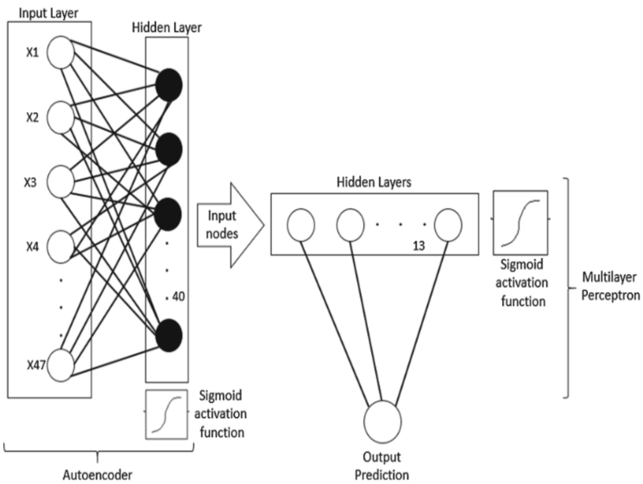


Fig 1: Auto-encoder and MLP Architecture

Methodology
Deep Neural Network

Artificial neural networks are used to carry out sophisticated computations in data-intensive deep learning methods. Graph theory is a type of machine learning that is based on human brain structure and function. Using deep learning methods, machines analyze examples to learn.

A Deep Neural Network relies on historical data to make forecasts, and until the boundaries and traits of the information are consistent, it makes no difference if the climate information is state-specific or local. Furthermore, state information can be simply thought of as the average of all local information. Henceforth, if a prototype shows excellent performance in the area of Jagatsinghpur in Odisha, then it will reveal similarly excellent performance for the whole state. Consequently, the factors are rescaled, avoiding scaling problems by using Minmax Scaler from the SciPy library for preprocessing with Sk-Learn. The dataset is then trained and tested.

In this study, the forecast expertise of the prototype is assessed through the utilization of a fold cross approval. An approximately equal number of folds [k=7] is considered, resulting in repeated training over the last six [k*1] folds. A portion of the fold is then used to determine its appearance. The information is then converted to a 1D vector structure using an entirely associated NN. As a result, a convolutional layer is added later. After that, the initiation functions are collected in a completely associated thick layer. In completely associated layers, there are significantly more neurons than in yield layers. Furthermore, the output layer is also an associated thick layer responsible for generating the expectations so that neural directions can think more broadly before joining the yield layer. The boundaries for the model must be established.

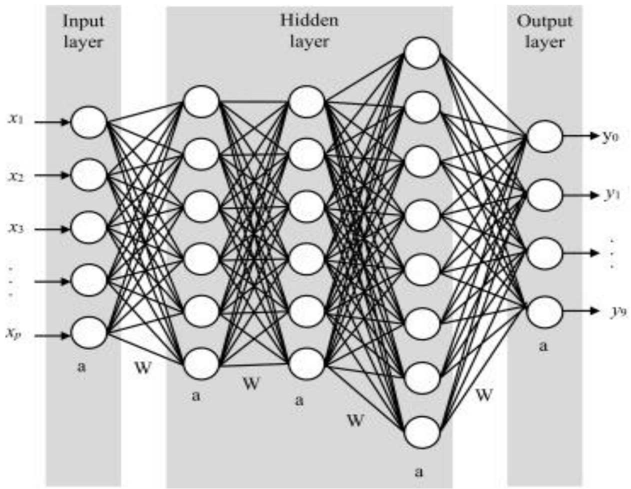


Fig 2: Deep Neural Network Architecture DNN

Project Modules
Data Visualization

To illustrate the practical approach to the classification of rainfall throughout everyday life, data visualization techniques are used. The practical way of categorizing rainfall is introduced through various charts and graphs. Before the introduction of data visualization techniques, large datasets were impossible to analyze and visualize There are two general categories of exploratory data analysis. Each method is maybe nongraphic or graphic, whereas secondly, each method is either multidimensional or unidimensional Data summarization that does not involve graphics is usually done through calculating summaries of statistics, while graphical methods represent the data through diagrams and pictures.

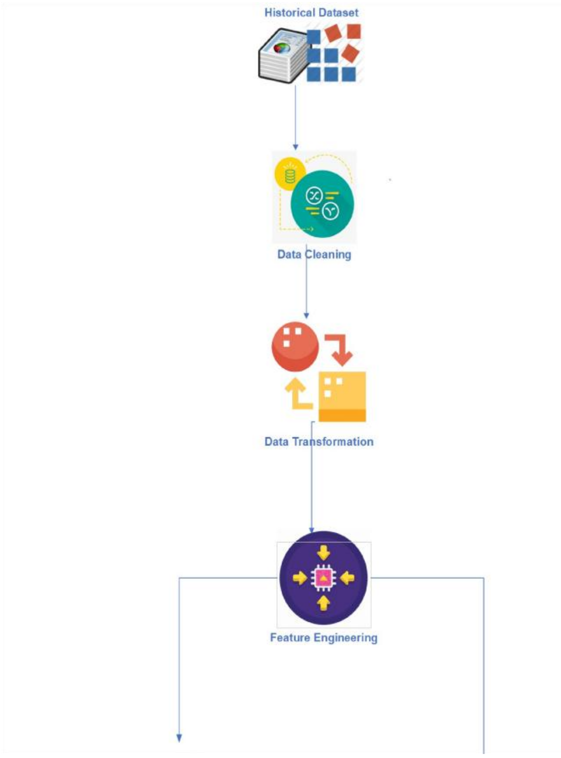
Preprocessing

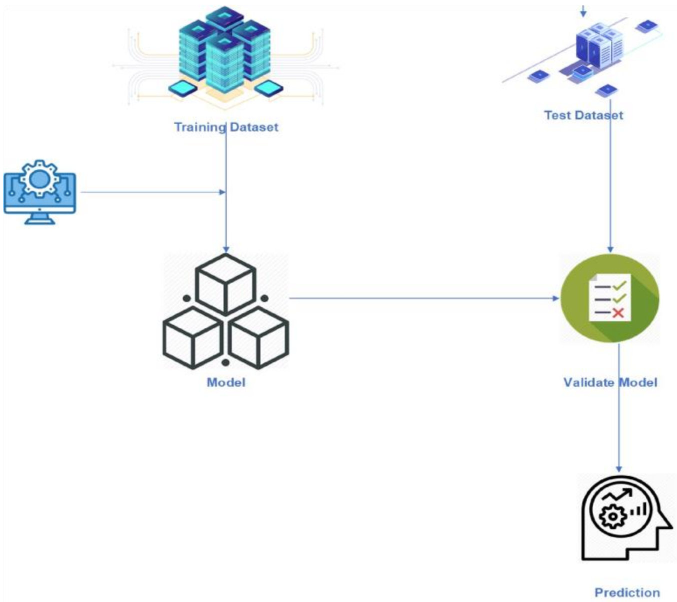
Indicator variables convert categorical data into boolean values by dividing the sample into intervals and replacing the categorical values. Pre-processing: Minor observation errors are reduced by pre-processing. Creating N-1 Columns: We need to create N-1 columns if we have more than two values (n). Centering and Scaling: The mean of one feature can be subtracted from all other features to calculate its center. This is used for scalability. Depending on the label's meaning (ordinal or nominal), one may represent an ordinal feature or not. The majority of labels possess a combination of both nominal and ordinal features. In what follows, you will discover how you could use the LabelEncoder class of sklearn to encode labels in numerical form for the learning algorithm.This module encodes labels for categorical features. Label encoding is the process of transforming word labels into numerical values, which algorithms understand and can manipulate accordingly.

Prediction

The parameters are passed into the DNN model as the first step. Random weights are used in this step. We recommend there be no more than one weight. With weight training, the desired outcome is to reduce neural network errors. This produces the desired outcome. In the case of incorrect output, weights on the DNN are adjusted to compensate. Therefore, resemblance rules the output. Then, test the weights against another dataset after adding the weights. The rainfall was predicted using deep learning techniques. Multilayer Perceptron and Auto-Encoders were used. A single neuron is present in the input layer of hidden layered neural networks(NN). Each layer is connected to this neuron. This process continues when the following data is incorrect. The sigmoid function is used to calculate hidden layers in neural networks. In Backpropagation NN, input signals are captured by input neurons, and output signals by output neurons.

Architecture Diagram





Obtaining the Past Dataset according to the requirements, cleaning the dataset (removing incompatible data, fixing missing data).It then goes on to simplify it through feature engineering. A better model can be trained if the past dataset is separated into train, test, and validation. A greater framework can be trained by utilizing the hybrid model. By evaluating the test dataset, the final output will be delivered.

Implementation and Results

A lot of experiments were conducted to find the proposed methodologies. For this study, we collected historical data from the Jagatsinghpur district of Odisha. Which contains various atmospheric parameters such as temperature, precip type, wind speed, wind speed, humidity, cloud cover, daily summary, etc.

	Formatted Date	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)	Daily Summary
0	2006-04-01 00:00:00+02:00	Partly Cloudy	rain	9.472222	7.388889	0.89	14.1197	251.0	15.8263	0.0	1015.13	Partly cloudy throughout the day.
1	2006-04-01 01:00:00+02:00	Partly Cloudy	rain	9.355556	7.227778	0.86	14.2646	259.0	15.8263	0.0	1015.63	Partly cloudy throughout the day.
2	2006-04-01 02:00:00+02:00	Mostly Cloudy	rain	9.377778	9.377778	0.89	3.9284	204.0	14.9569	0.0	1015.94	Partly cloudy throughout the day.
3	2006-04-01 03:00:00+02:00	Partly Cloudy	rain	8.288889	5.944444	0.83	14.1036	269.0	15.8263	0.0	1016.41	Partly cloudy throughout the day.
4	2006-04-01 04:00:00+02:00	Mostly Cloudy	rain	8.755556	6.977778	0.83	11.0446	259.0	15.8263	0.0	1016.51	Partly cloudy throughout the

Fig 3: Past Dataset contains atmospheric parameters

Errors Calculation

Approaches are rated in terms of their accuracy against types of errors that may negatively impact the algorithm. MAE, MSE, and R-square are calculated. Using MAE, all absolute errors are calculated as a mean value. A mean value is first calculated for all the datasets present, then the mean value is subtracted from each dataset individually and added together.

$$MAE= 1/n\sum|xi-x|$$

Similar to mean absolute error, MSE measures mean standard error.

$$MSE=1/n\sum(xi-x)^2$$

Our total amount will be stored as a value. This value will then be divided by all the current values. The result is then squared.

$$R^2 = 1-x/y$$

Visualization

Graphs are used to represent the dataset. After the process is complete, different graph types are generated. The types are listed in the following Plots.

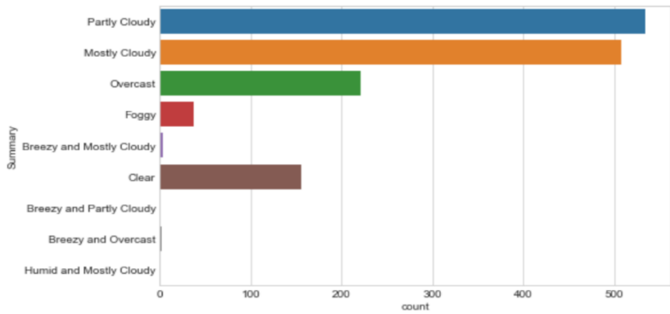


Fig4: Parameters in Summery column

Fig5: Heatmap of the dataset

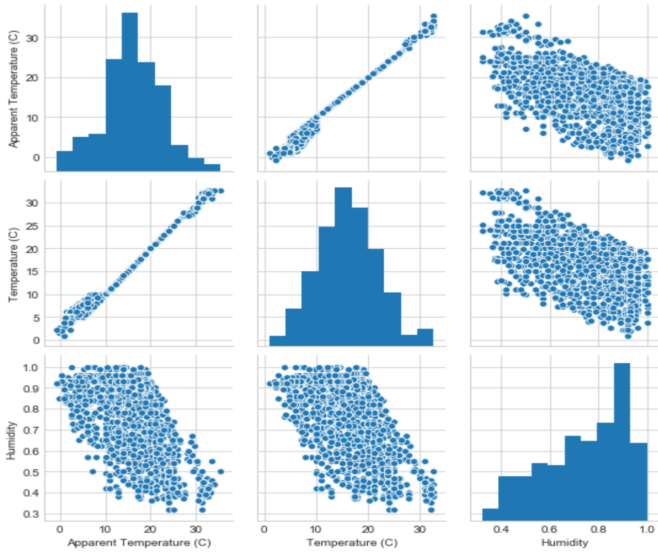


Fig6: Pair plot Temperature, Apparent Temperature, and Humidity

To predict the rainfall effectively, we can utilize Machine_Learning (ML) and Deep_Learning (DL) models. The Deep Neural Network (DNN) helps the historical data to convert into neurons to train the proposed model. We separated the data into train and testing and calculate the Mean Absolute Error (MAE) after DNN training. We calculate and analyze the number of iterations to determine which approach is best. When a prediction is placed in a straight line with slope 1 and intercepts 0, one can consider it perfect. A straight line tends to follow the indicated straight line when data points are plotted. Specifically, we found that the effect was strongest on rainy days, which was the focus of our study. Our study is less significant on days with the least rainfall i.e less than 20mm1, which tends to deviate from the target value for days with light rainfall. We will aim for improving the predictions in the future for such cases, and this indicates that there is room for improvement in the proposed architecture.

Conclusion

The observed rainfall and precipitation predictions agree more closely than the observed rainfall and simulated precipitation magnitude. As a result, we can consider the hybrid DNN as a potential method for better rainfall prediction. Our research utilized historical rainfall and flood flow data to develop a different model: the use of rainfall-flow patterns that predict short-term flood flows. Hydrological features are extracted and spatial-temporal metrics for rainfall patterns are used in our models in real-time to predict flood processes. These experimental results based on datasets from various models have shown that the proposed models are exceptionally effective at predicting rains in real-time in wet and drought-ridden watersheds. In this paper, various types of approaches and approaches to forecasting rainfall are discussed along with potential issues found by applying these different approaches. Artificial Neural Networks are a superior solution to all other approaches because they can learn from the past and are not linearly related to rainfall datasets.

Future Works

Plans for improving our architecture will focus on light rain scenarios. As the scenario that is most likely to lead to negative consequences is heavy rain, our research was precisely centered on the opposite, heavy rain scenarios.

References

[1] M. D. Crown, "Validation of the NOAA space weather prediction center's solar flare forecasting look-up table and forecaster-issued probabilities," Space Weather, vol. 10, no. 6, pp. 1–4, 2012.

[2] C. W. Zheng, C. Y. Li, X. Chen, and J. Pan, "Numerical forecasting experiment of the wave energy resource in the China sea," vol. 2016, no. 3, pp. 1–12, 2016.

[3] K. C. Luk, J. E. Ball, and A. Sharma, "An application of artificial neural networks for rainfall forecasting," Ma

[4] K. W. Appel, R. C. Gilliam, N. Davis, A. Zubrow, and S. C. Howard, "Overview of the atmospheric model evaluation tool (MET) v1.1 for evaluating meteorological and air quality models," Environmental Modelling & Software, vol. 26, no. 4, pp. 434–443, 2011.

[5] P. Montero-Manso, G. Athanasopoulos, R. J. Hyndman, and T. S. Talagala, FFORMA: Feature-based forecast model averaging," Int. J. Forecasting, vol. 36, no. 1, pp. 86–92, Jan. 2020, doi: 10.1016/j.ijforecast.2019.02.011.

[6] Grace, R. K., & Suganya, B. (2020). *Machine Learning-based Rainfall Prediction. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS).* doi:10.1109/icaccs48705.2020.9074233 10.1109/icaccs48705.2020.9074233

[7] Balamurugan, K., Uthayakumar, M., Sankar, S., Hareesh, U. S., & Warrier, K. G. K. (2018). Effect of abrasive waterjet machining on LaPO4/Y2O3

ceramic matrix composite. *Journal of the Australian Ceramic Society*, *54*(2), 205-214.

[8] Kala, A., & Vaidyanathan, S. G. (2018). *Prediction of Rainfall Using Artificial Neural Network. 2018 International Conference on Inventive Research in Computing Applications (CIRCA).* doi:10.1109/icirca.2018.8597421

[9] Ramesh, S.S.S., Rathore, K.S., Raj, R., Vatsalya, K., Vatsa, M. Integrated malware analysis using markov based model in machine learning, International Journal of Engineering and Advanced Technology, 2019, 8(4), pp. 219–222.

[10] L. F. Richardson. Weather prediction by numerical process. Cambridge University Press, 2007.

[11] Subashka Ramesh, S.S., Hassan, N., Khandelwal, A., Kaustoob, R., Gupta, S. Analytics and machine learning approaches to generate insights for different sports, International Journal of Recent Technology and Engineering, 2019, 7(6), pp. 1612–1617.

[12] Minu, M.S., Aroul Canessane, R. Secure imagetransmission scheme in unmanned aerial vehicles using multiple share creation with optimal elliptic curve cryptography, Indian Journal of Computer Science and Engineeringthis link is disabled, 2021, 12(1), pp. 129–134

[13] Minu, M.S., Arun, K., Tiwari, A., Rampuria, P.,Face recognition system based on haar cascade classifier, International Journal of Advanced Science and Technology, 2020, 29(5), pp. 3799–3805

[14] K, Balamurugan, M, Uthayakumar, S, Sankar, US, Hareesh, & KGK, Warrier. Predicting correlations in abrasive waterjet cutting parameters of Lanthanum phosphate/Yttria composite by response surface methodology. Measurement, 131, 309-318, 2019

[15] Latchoumi, T. P., & Parthiban, L. (2021). Quasi oppositional dragonfly algorithm for load balancing in a cloud computing environment. Wireless Personal Communications, 1-18.

[16] Pavan, V. M., Balamurugan, K., & Latchoumi, T. P. (2022). PLA-Cu reinforced composite filament: Preparation and flexural property printed at different machining conditions. Advanced composite materials. 31(1), 102-117

[17] Banu, J. F., Muneeshwari, P., Raja, K., Suresh, S., Latchoumi, T. P., & Deepan, S. (2022, January). Ontology-Based Image Retrieval by Utilizing Model Annotations and Content. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 300-305). IEEE.

[18] Latchoumi, T. P., Swathi, R., Vidyasri, P., & Balamurugan, K. (2022, March). Develop New Algorithm To Improve Safety On WMSN In Health Disease Monitoring. In 2022 International Mobile and Embedded Technology Conference (MECON) (pp. 357-362). IEEE.

[19] N. I. Sapankevych and R. Sankar. Time series prediction using support vector machines: a survey. Computational Intelligence Magazine, IEEE, 4(2):24–38, 2009.