MVI and Forecast Precision Upgrade of Time Series Precipitation Information for Ubiquitous Computing

Ashok Kumar Tripathi[1], P. K. Gupta[2], Hemraj Saini[3], Geetanjali Rathee[4]
[1,2]Department of Computer Science & Engineering, Jaypee University of Information Technology, Waknaghat-173234
[3]School of Computing, DIT University, Dehradun-248009
[4]Department of Computer Science & Engineering, Netaji Subhash University, New Delhi
E-mail: ashu_jhansi@rediffmail.com, pkgupta@ieee.org, hemraj1977@yahoo.co.in, geetanjali.rathee123@gmail.com

Keywords: stochastic gradient descent, RSM-Prop optimizer, ADAM optimizer, LSTM, missing value imputation, biquitous computing, extended kalman filtering

Received: May 4, 2022

Missing values are the major problem in a time - series dataset that led to a very typical phenomenon in a dataset analysis. Better analysis of a dataset in ubiquitous computing can only be achieved after efficiently handling the missing values. This paper envisioned to present the missing values imputation (MVI) in the monthly rainfall dataset in India. Generally, the approximation of the missing values was completed using Non-linear principal component analysis, which is having a scope of enhancement. Kalman filter using Arima model to impute missing values is a better way that can be further improved by using Extended Kalman filtering. Further, a rainfall prediction is carried out using LSTM along with three different optimizers including stochastic gradient descent (SGD), RSM-Prop, and ADAM optimizers. In addition, a comparative prediction of rainfall is depicted which shows the combination of Extended Kalman imputation, LSTM, and ADAM optimizer to outperformance in prediction. Extended Kalman Filter (EKF), which was the primarily industrialized for the Apollo Mission. The EKF is built for the predication that is the linearization of the structure.in this research proposes for enhanced extended kalman filter for missing values imputation. The outcomes display that the proposed EKF with the data imputation is able to precisely estimation the rain fall compactness even when data is not available, and it help for weather prediction.

Povzetek: Predlagana je izboljšana metoda za nadomeščanje manjkajočih vrednosti v časovnih vrstah.

1 Introduction

India is having various kinds of climate that includes all types' weather in the world, but the weather in most of the nation is truly hot and affected by the rainy season regime, with a dry and a monsoon season. The rainfall, much or less penetrating and long-lasting is depending upon the locations, where the rainy season period is from the month of June to September excluding in the southeast, where the retreating rainy season continues till the end of the year. Rain water released from agricultural and water system activities is required at various time scales. Due to irregular rainfall, agriculture and the regular works are very much affected in our country. The growing rate of the food products is less, that's why the supply and demand is also affected [1]. Rainfall is very important for agriculture, resource management for water, management of food and for transportations for ensuring an efficient ubiquitous computing in the network. Heavy rainfall might cause flooding, landslides and additional hydrological disasters that interrupt the humanoid life activities, common economy, and the environment. We will suggest improving the agricultural growth to analyze ubiquitous computing and forecast the time series of Indian rainfall.

These missing values are generating many flaws in the dataset like communication failure, lack of information, signal problem, records not maintained properly, failure of calculating gadget, faults in observation recording, , climatologically immoderations, failure the network produced by disasters and the challenges related to accessing dimension zones.[2] Henceforth, it is essential to apply a suitable filling of missing data of any previous study. Likewise, this missing information may decrease the accuracy of the determining insights on the grounds that there are fewer facts than originally arranged. In addition, the worry is that the presumptions behind the numerous assessment methods depend on the whole cases and the missing values can confound the hypothesis required.

These are the main concerns of hydrological investigators, that may be considered by the eroticism in rainfall and recurrent rainstorms universally, through incessant rainfall data input, which positively characterizes the indispensable portion of hydrology and water superiority study. Characteristically, rainfall data are got from ground location and the remote sensing organisms. At the similar time, rainfall location frequently serves as the rudimentary basis of extended rainfall data. Howsoever smooth the finest apparatus is used; investigators frequently face data shortage quandaries owing to electromagnetic fulmination pulses and bring down the performance of the device as well as the properties of human error [3]. This data shortage quandary can cause enormous difficulties in procurement continual rainfall data sequence and influence hydrological and aquatic excellence examinations, which are to be sure significant for water resource management. There are several reasons for the missing value imputation and predication of rainfall. There are many types of problems faced due to these missing data [4]. Missing data weakens the quality of data and also effect when analyzing the prediction base on this type of dataset. This problem has controlled a wide - range of research on emerging the approaches for the imputation of missing data. The imputation technique re-establishes the missing values in the dataset for analysis features rules and relationship. There are many studies conducted for rainfall missing values imputation and the prediction that uses many types of imputation and prediction models and methods, by describing some techniques which uses for missing values imputation.

1.1 Problem statement

There are some imputation techniques which are used for the time series data set. One of the most popular techniques which are used for handling the missing values is mean; in the mean approach we calculate the mean of record and fill the values in place of messiness, the next way to handling missing values by forward filling and backward filling are two ways to deal with missing values.[5] Forward filling implies fill missing qualities with past information and backward, filling implies fill missing values with the next information point [6]. In the existing method of missing values imputation have some problem, so it does not give the close and very accurate results. In this technique there are no associations between features, it works in an only singular direction on column level. The existing techniques have to stipulate the columns that hold information nearby target column which will be imputed. These techniques will also provide poor outcome on pre-arranged categorical structures. It does not have an explanation for the doubts in the imputations [7].

1.2 Research objective

In this paper we are using a new technique for the missing values imputation called Kalman filter and extended data and analysis which technique give the close result for ensuring an efficient computation. The best technique applies for prediction of Kalman filter [8]. Both techniques are applied to the Indian rainfall dataset for handling the missing rainfall in India. Here we will apply the LSTM algorithm which is a machine learning algorithm also using the three different optimizers with LSTM model and compare with the optimizer that provides the better and close result to find the best approach for filling the missing value gaps [9]. The consequences are designated that the rainfall dataset shortage during low flow times would outcome in inferior model presentation and superior forecast uncertainty, particularly to the certain smallest value, and time once the extreme values are additionally vulnerable, is the rainfall data shortage throughout the large flow periods and the preparation of rainfall data and the model presentation got by the LSTM procedure is the greater to the conventional algorithm and climate producer presentations. This is benefit of the Extended Kalman procedure would be additional conclusive if a specific beginning of facts shortage is reached. It is well-known that if the finest algorithm is used, the credited value is continuously inferior compared to the top experiential values. In this research, we focused on Extended Kalman Filter and Kalman filter-based models and effort to reformulate the way Kalman filter is used for missing values imputation rainfall imputation models, even when the input rainfall data covers missing values. We have proposed a unidirectional LSTM technique for ubiquitous computing architecture for network-wide rainfall state prediction to address the above-mentioned failings.

1.3 Contribution

The assessment of the forecast competence of LSTM based models has the potential to facilitate the further research on the deep learning method projected for rainfall prediction difficulties. Experiments based on one real-world datasets with different missing value specify that the proposed architecture that can achieve outstanding imputation and prediction results. In summary, our contributions can be addressed as follows:

1. We have suggested Kalman filter and Extended Kalman filter structure with an imputation component, to fill the missing data in the time-based input data and in reappearance to help recover prediction accurateness for ubiquitous computing. In this research the researcher is try to find rate of change between Measured Value and Actual Value which is predicted with kalman filter and try to find more accurate result with derivative value which is implemented in extended kalman filter. We have suggested a loaded bidirectional and unidirectional LSTM architecture. LSTM, for network-wide rainfall forecasting. This stacked planning with multiple layers is flexible. The assessment of the forecasting ability of LSTM based models has countless possible to facilitate the strategy of neural network technique for rainfall prediction. A practical rainfall state data is tested in this research where the dataset is available on Data-world.

The rest of the text is organized in different steps as mentioned further. Details of data collection, dataset strictures, and missing values investigation in the time period of interest are defined in Section 2. Connected methods for the data assessment are described in Section 3. In Section 4 describes the introduction of Kalman filters. The proposed approaches, solution and implementation of the Kalman Filter are defined in Section 5. In Section 6, we will present the results of our comparative study; examine the overhead of the Kalman Filter operating as an amount of the service time; and the forecast accurateness of Kalman filters. Conclusion and future work are discussed in Section 7.

1.4 Missing value imputation and prediction of rainfall as a case study

1.4.1 Data

The Indian Administration has assumed numerous investigation studies for examining the influence of worldwide warming. There are many changes in climate and rainfall scenario in India. The existing structure of the India National Disaster Risk Reduction and Management (NDRRMC) uses climate positions to deliver weather associated information and facilities make use of the rainfall calculating appliance and sensors. The automated weather station (AWS) organization that events. The number of water, predictable in a selected month and place doesn't have a good rainfall predicting model for calculation of how much rainfall produces in the months. Historic rainfall data from the given weather state can be used to produce a rainfall predicting model which comprises the records of impressive data occupied from the measurement sensors and campaigns.

Here is a well-known propensity of the growing occurrence of extreme rainfall (heavy rainfall proceedings) over India, especially over the relevant elements of India for the duration of the June to September is the rainy season period. Here isn't any evidence of world warming at the determined adjustments in yearly or periodic rainfall over India. However, there may be developing evidence signifying that the growing incidence of excessive rainfall is because of worldwide warming. The weather alternate evaluation made via way of means of the Intergovernmental Panel on Climate Change (IPCC) advises that during the future, the frequency of intense rainfall can also boom over India because of the boom in worldwide warming. Though, here are no different long-time adjustments/tendencies in rainfall over India, which may be attributed to worldwide warming. The Indian Monsoon is determined to be a solid system. This information with extraversions of common rainfall, its miles very hard for a statistical model to be expecting the specified information point. Here we put in force neural networks too are expecting the average rainfall, the neural internet is used to create more than one function that enables predicting the information factors with extra seasonal versions. We are providing below, the graphical format of the monthly rainfall viewed with the units in Millimeter in every State of India.

The Indian government had made rainfall dataset for experimental studies. We are using the rainfall dataset for research which is available on the Data - world and Kaggle. There is added 3000 rain-gauge positions feast all over the country for the 115 year dataset from 1901 to 2015. The principal implications of those studies based on the 115 years of rainfall dataset. Here we can view the Indian rainfall through graphical representation. In this study we observe that how much rain occurred in a particular month in a particular state. In this graph, we are displaying rainfall in a month with state and years in X axis and Y axis display how much rainfall occurred at month.

The most ideal technique for taking the missing information is to forestall the issue by well-arranging the review and gathering the information cautiously. Coming up next are proposed to limit the measure of missing information in the dataset for a better outcome and examination. Missing information, may truly compromise implications about from randomized experimental preliminaries, particularly if missing information is not dealt with appropriately as depicted in Figure 1. In this paper the research is based on missing values imputation and prediction of as per the rainfall information which is given in the dataset. There is analysis how much missing information in the dataset.

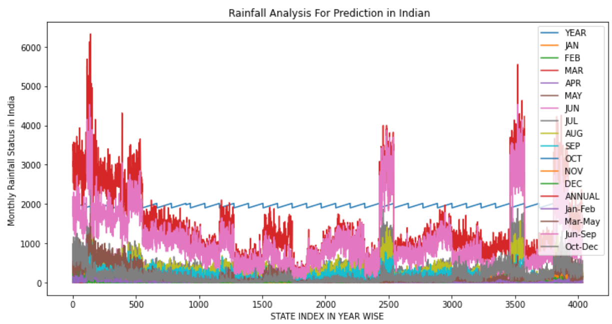

Figure 1: Monthly rainfall status in India.

This plot characterizes the association between the missing values by column. Here we are using heat map, which will assist us understand how the fields relate to each other in regards to soreness (incompleteness). Even though we can visualize that relationship by looking at the matrix, heat maps make them more evident. It's not always that we'll have our rows sorted in a convenient way for us to realize the pattern. Here we display the monthly missing values in the column, in this graph we represent in which month how much information is missing in the dataset.

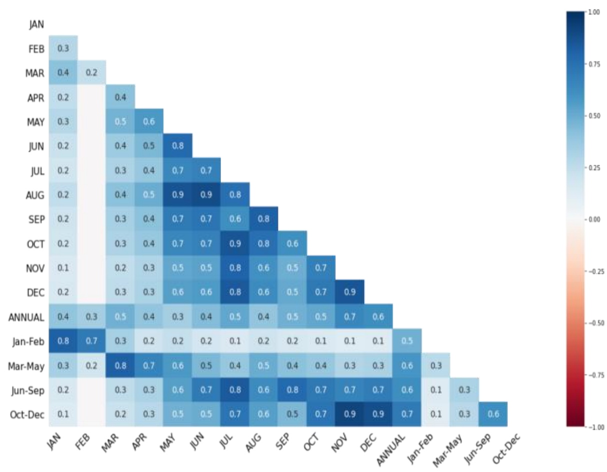

Figure 2: Display missing values in India rainfall dataset.

Previously here are numerous studies existing on the experiential trends and the inconsistency of the rainfall and also dangerous rainfall proceedings, but all the revisions are grounded in the last 115 years and more statistics and the current years are not included in the research. The current reports, all the study of experiential rainfall patterns, trends and inconsistency had been completed based on current beyond 35 years (1981-2015) that will assist to have the influence of the current variations for climate change adaptation and organization by the state establishments as depicted in Figure 2. Daily Rainfall records from 1981 to 2015 are considered for evaluation in our research. From the daily rainfall data, month-to-month rainfall series of every 12 months are computed and then monthly rainfall collection has been constructed by considering the arithmetic average of all of the month rainfall values in the state. The month-to-month rainfall series of the state has been computed by the use of region-weighted rainfall values of all of the districts in the state. Due to not proper visualization of result we fix our research on a state basis, here we are selecting only one state which is Bihar and all methods apply for research.

In these studies, we are focused on what's best way apply for imputation of missing values and predicting the rainfall for growth which depending on rain. This weather state has had apparatus for calculating impressive circumstances, this will deliver result makers thru climate associated decisions. This has recognized that, these states implement well in gathering climate data with satisfactory stages of accurateness. Due to the large amount of dataset the graphical representation is very typical in that reason we are focused only one state of visualization for the clear records.

By this, a forecast technique specifically a month rainfall forecasting model using LSTM with optimizer is projected in the research. [13] This research objective to focus on proposing an efficient rainfall imputation technique and forecasting technique by showing rainfall data groundwork as well as applying and assessing a Multilayer Perceptron Neural Network model.

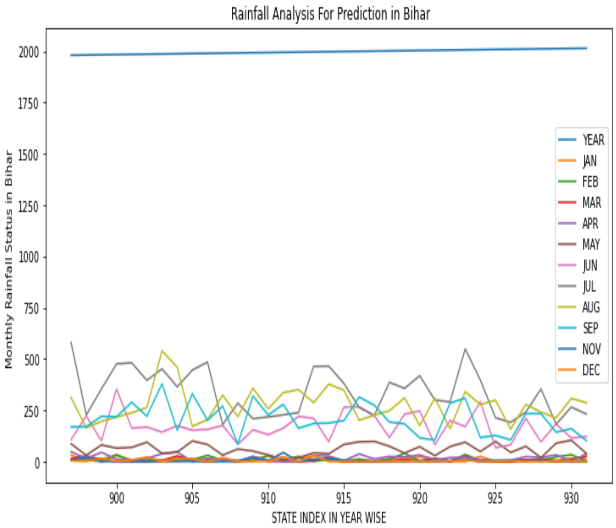

Figure 3: Rainfall analysis for Bihar.

The outcome of the research will deliver significant the method involved with settling on decisions by recognizing a choice, gathering data, and surveying elective goals. Through this research it is expected to contribute to the current knowledge of rainfall predicting by developing a rainfall missing values imputation and prediction model that might benefit climate stations deliver better rainfall info to particular locations. In the dataset, missing information, or missing values, happens when no data information is put away from the variable in an observation. Missing information is a common occurrence and can significantly affect the conclusions that can be drawn from the information as depicted in Figure 3. A time series database (TSDB) is improved to sort and establish the information computed by time date and year [11]. A period series is a collection of information focuses that are assembled at progressive stretches and recorded in time request. Time series information bases are especially helpful to monitor the access metrics, disappointment metrics, process behavior and responsibility checking. TSDBs can figure out enormous and complex measures of information, making the data more open than if it were put away in a conventional data set. Here are some differences between regular data and time series data [12].

The TSDB includes a time and date field, Time series data too discloses an all filled of a system over time and allows the study of historical tendencies. Time series information is constantly gathered throughout a predefined time-frame and information from responsibilities is new and composed as additions, instead of refreshed to supplant the information that as of now exists. At the point when information is composed, it is consequently allocated to the latest time span.

2 Literature review

The missing information mechanism has a huge impact on the dataset. Here we analyze about how many types of missing values are there and how to deal with these missing values, On that basis the results of the study to recognize its category in order to select the most suitable technique to deal with missing information. Absence of information in the time series dataset can lead to prejudice outcomes and can sometimes totally prevent significant studies to be accepted. Here are the different kinds of circumstances when the information is missing. There are mainly three categories of data which is dealing with the missing value. These mechanisms are missing completely at random (MCAR), missing at random (MAR) and missing not at Random (MNAR)

2.1 Missing completely at random

Go with the name missing completely at random here it is independent of any object an unobserved variable in your dataset, there is the probability of the missing data is not related with any other variable. It is not depending on any other variable in the dataset, and there is no relationship that you find missing values and data point that have in your dataset. That is why less data for statistical calculation is available and researcher gets reduced data set.

2.2 Missing at random

Missing data within the subgroup of that data and it is affecting research's other variable also, there is some dependency between variable. Missing at random means that the tendency for a data point to be missing is not related to the missing data, but it is connected to the particular of the observed data set.

2.3 Missing not at random (MNAR)

Data missing is systematically related to observed data, that means this type of missing information can be related to the event and any other factor which is no not been captured by the researcher. The researcher has missed the data and instrument have not captured this data, because of various reasons. Lacking little information could have got but have not got because of some things is not estimated. There are two possible reasons are that the missing value depends on the hypothetical value. For example: people with high salaries generally do not want to reveal their incomes in surveys.

Deep learning Technique, by way of a subdivision of machine learning technique, develops popular and quickly be accepted in the rainfall forecasting zone. Maximum of the a new projected rainfall forecasting technique is based on neural networks [9], which has mostly processed sequence information by maintaining the structure and internal memory. To discourse explosion gradient difficulties, Long Short-Term Memory network (LSTM) is designed to learn long-term dependencies of preparation data via a gate and memory components. Several current studies accepted the LSTM as a base point or structure blocks in their projected technique for rainfall predication. Though RNN and its alternatives have been accepted as building blocks of rain prediction technique, some researcher re-formulated their model structure to recover rainfall prediction accurateness and sturdiness. There are many type missing data imputation technique for the time series dataset [14], including regression, spectral analysis Method and, EM procedure have been implemented and applied to estimation for the missing values [15]. Among these are non-deep learning-based models, matrix factorization technique generally can achieve advanced forecast accurateness. A new Bayesian time-based matrix factorization technique is planned to resolve spatiotemporal forecasting problems once here is the missing values in input data [16]. This technique can deal through the missing data and reach good forecasting accurateness. However, the mechanisms of these approaches and the sizes of the data used to train models are importantly dissimilar from those of deep learning-based method [17]. Also, merging these data imputation techniques with rainfall prediction method frequently leads to the two-step procedure.

For the basis of the prediction problem with LSTM technique, the contribution of time series contains lost data or null data, the method will fail owing to missing data cannot to be calculated throughout the training procedure. If missing values are usually as selected pre-defined values, like mean, zeroes of historical observation, or previous detected values, there is biased ideal inputs will effect in biased parameter approximation in the preparing process [18]. Additional, resolving the prediction and imputation tasks at the similar time frequently outcomes are detached with imputation and prediction technique [19].

Kalman filtering, defend as Kalman smoothing on the state presentation of an ARIMA technique, is frequently measured a good method for the imputation of a very seasonal dataset. The imputation procedure of Kalman filtering is also used imputation. Kalman filtering is not able to deliver a sensible imputation for the variable regular cell throughput owing to the huge times of the missing values [20]. This kind of method produces numerous imputed values for every missing field, and these imputed data are joint to generate a quantity of whole datasets. This procedure is repeated multi times for processing of each data set, which leads to several types of outcomes that are used to estimate the target variable and approximation its accuracy. Here the procedure is separated into three phases: imputation, analysis, and prediction [21].

3 Kalman filter

Rudolf E. Kalman was published the object for an Innovative Method for the Linear Filtering and Forecasting Complications, Kalman (1960), anywhere he proposed the explanation of the traditional filtering and forecast difficulties using state space representation for the random procedures. [22] The technique projected in the future, be known as the Kalman filter. The Kalman filter is broadly utilized in the field of control engineering, signal handling, the measurable control of value and for the missing values imputation. [23] There is a ton of sources that is giving great introductions with the Kalman filter. It is exceptionally normal for a lot of information to be absent in natural information in time series data set. [24]The Kalman filter empowers the handling of time series with missing information. In this research, we will apply the Kalman filter to estimations of wave statures where about some amount of the values is absent. There have been different works handling the execution of the Kalman filter for dataset with missing values. We try to find what a shortcoming and what are the strengths of the Kalman filter that develop apparent and deliberate the difficulties that happened in the implementation of Kalman filter [25, 26].

4 Extended Kalman filter

The extended Kalman filter is the nonlinear type of the Kalman filter which is learners about an estimation of the existing mean and covariance. In the case of well-defined evolution techniques, the Extended Kalman filter has been considered the de-facto standard is the concept of nonlinear state assessment. The Extended Kalman filter, modified techniques of calculus, specifically multivariate of series developer of the luminary's technique around a working position. If the model is not well known or is inaccurate, then Monte Carlo technique, especially particle filters, is active for estimation. Monte Carlo techniques precede the presence of the Extended Kalman filter are more computationally expensive for any moderately dimensioned state-space. The progress of EKF, which decreases the effects of the difficulties which is arising in nonlinear systems afterward conducting a linearization finished the greatest actual state approximation

5 Proposed mechanism

In this research, the unusual rainfall datasets remained intercalated with the help of two missing values imputation approach, and after then we are comparing with the imputed values and find value is used to recognize the effects of two imputation technique on the rainfall datasets. These interposed rainfall data sets were formerly used as an input for measuring the choice of dissimilar imputation technique for simulations. The complete procedure of this study is exposed in figure 1.

In the part of this section, first we deliberate the methodology of imputation as well-known imputation techniques with the process of application to the Rainfall dataset. These approaches contain the univariate Kalman filter and Extended Kalman filter. Kalman filter, also known as Kalman smoother. Here we use two models with Kalman filter its representation of as ARIMA model and StructTS model. Is frequently measured a good technique for the imputation of an extremely seasonal univariate dataset.

This investigation thinks about the Extend Kalman filter based way to deal with reservoir sampling and histogram-based methodologies. We show that Extend Kalman filter and Kalman filter is imputed the missing values and has the least pull mean squared error for most cases. We present a novel explanation for embedding this estimation procedure. We describe the issue of irregular missing values in the sensor and device streams, and suggest a technique dependent on Kalman filter and Extend Kalman filter. For demonstrating the information sensor streams and instrument streams as a period series and utilizing Extend Kalman filter to predict and impute the missing values. We limit our extension in this paper of univariate time series, including of a solitary (scalar) perception recorded consecutively throughout equivalent time increases. In "multivariate time series" every time series perception is a vector of numbers. The methodology depicted here can be reached out of multivariate time series utilizing the particular technique. An Extend Kalman filter is an ideal recursive information handling and numerical assessment calculation that has been utilized for information osmosis and missing values predication. The Extend Kalman filter consolidates verifiable data for estimation of the current worth of the factors of interest. A Kalman filter can be introduced to various state techniques like remarkable technique and dynamic linear technique. My propose methods be dynamic linear techniques as it is the use of simple, takes some strictures to be initialized, and dynamically updates its state. [27] The dynamic linear technique is the difference in values of proceedings over time. With the help of Extend Kalman filters, this technique has enabled actual one pass forecast receiving values. In the presence of values in a stream, the Extended Kalman filter continuously forecasts and updates researcher with the successive data. Here we are given the implemented algorithm of Extended Kalman filter which is used for missing values imputation [28].

Algorithm:

```
def hx(x):
"Compute measurement for slant range that would correspond to state x”.
k = np.append((x[:,1][1:]-x[:,1][:-1])/(x[:,0][1:]-x[:,0][:-1]), 1)
y = k.mean()
if np.isfinite(y):
k = (k/y)
k[np.isneginf(k)] = -1
k[np.isinf(k)] = 1
k[np.isnan(k)] = 0
k = np.concatenate(
(k.reshape((x.shape[0], 1)), np.ones((x.shape[0], 1))), axis=1).
reshape(x.shape)
return x*k
```

Flow Chart of Proposed Algorithm:

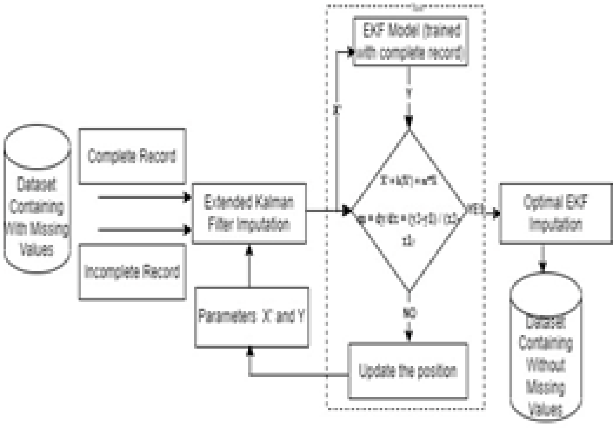

Figure 4: Process of extended Kalman for missing value imputation.

In figure 4 the flow diagram researcher describes how the algorithm flows the program hears we take a dataset with missing values and apply the Extended Kalman filter. There are two parameters X’ and Y and the equation applies Y= (hX’). Here is the h function that specifies how our location is mapped to polar coordinates x′ is used for predicted values and y is used for the difference between the measured value and actual value.

5.1 Function h(X′)

This is a purpose that specifies the mapping between our predicted values in Cartesian coordinates and Polar coordinates. This mapping is required because we are predicting in Cartesian coordinates, but our measurement that is coming from the sensor is in Polar Coordinates.

The Kalman filter functions to find optimum estimates and is expected to be Normal so what the Kalman filter actually does is to calculate the conditional mean and modification of the distribution for conditional on observations up to time. This research proposes is improved Extended Kalman Filter (EKF) with an adaptive structure and take close derivative result and multiply with rate of change in extended kalman filter.

6 Result analysis

This section explains various graphs that are used to determine the outperformance of the proposed mechanism over various existing approaches. The below Figure 5, is based on 6 years (2007 to 2012) observed records of rainfall data in Bihar state. In X axis we display how much rainfall occurred and measured in millimeter and Y axis we are displaying monthly rain produce.

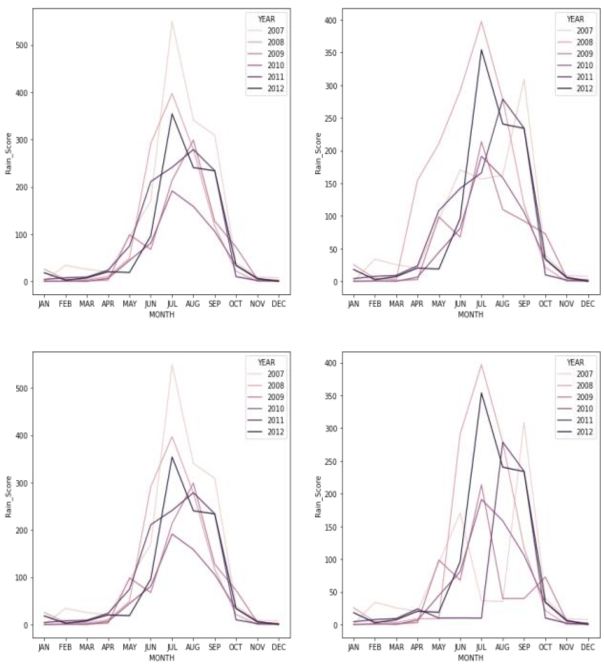

Figure 5: Rain score for Bihar 2007-2012.

The Figure 6 is displaying the comparison of result analysis in graphical form with the missing values imputation with original values, Kalman filter values and our propose algorithm extended Kalman filter values. The imputed result with original values and that is being displayed in blue color and predicted values are being by Kalman filter represent and displayed in orange color and predicted values are being by extended Kalman filter represent and displayed in green color.

In addition, the depicted Figure 7 compares the original values with stochastic gradient descent (SGD) optimizer. When the result is found the propose algorithm try to find better result by the use of Stochastic gradient descent (SGD) optimizer. In addition, in Figure 8 SGD resolved the Gradient Descent problem by using only solitary records to update strictures. But, still, SGD is slow to converge since it wants to advance and retrograde propagation for each record and the path to spread global minima develops very deafening.

Gradient Descent is the prevalent optimization method in Machine Learning and Deep Learning, and this can be used with most, if not all, of the learning procedures. The gradient is a slope of a function that measures the degree of alteration of a variable in reaction to the vicissitudes of additional variable. Mathematically, Gradient Descent is a curved function whose outcome is the incomplete copied a set of strictures of his contributions.

When a researcher has not found the best result with stochastic gradient descent (SGD) optimizer we try to change the optimizer and replace with RSMProp optimizer. After execution, we find the best result as compared to SGD optimizer. Below we can view in the graph, the comparison of the original values with its accuracy is better than the previous optimizer.

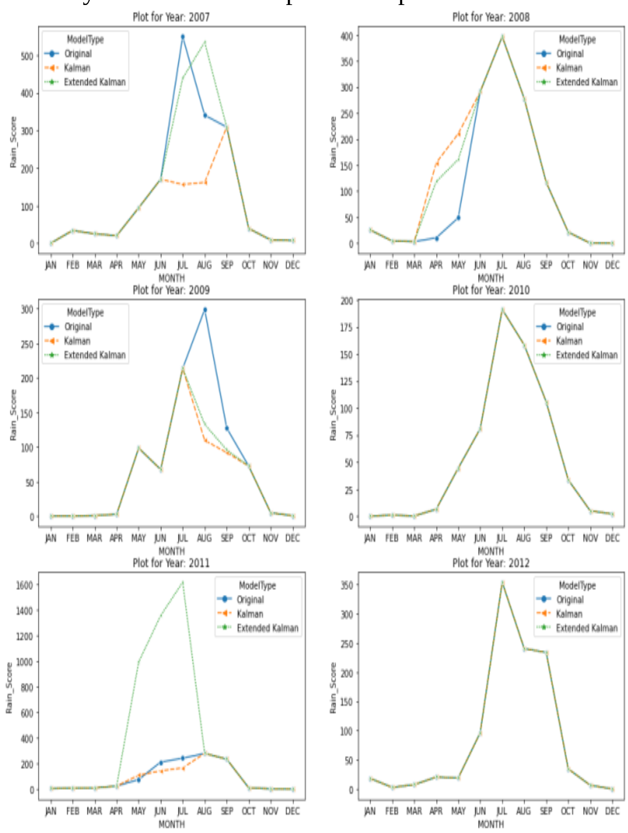

Figure 6: Missing values imputation comparative analysis with original, Kalman filter and extended Kalman filter.

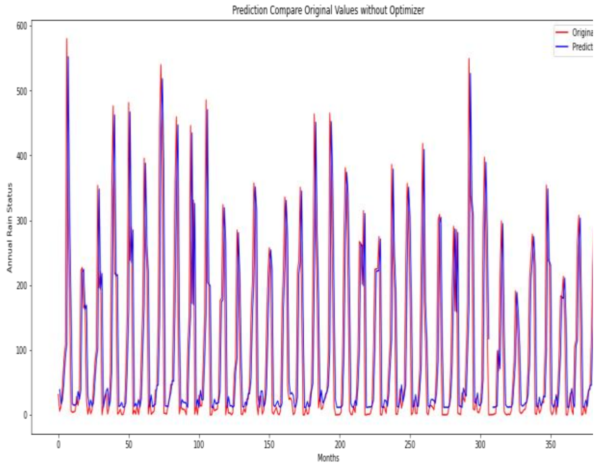

Figure 7: Predicted values compare with the original values.

The RMSprop optimizer is comparable to the gradient descent procedure with impetus as depicted in Figure 9. The RMSprop optimizer limits the fluctuations in perpendicular direction. Consequently, we can upsurge our learning rate and our procedure might take superior ladders in the horizontal way converging earlier. There is alteration between RMSprop and gradient descent is on how the gradients are computed. There is subsequent computation show how the gradients are computed for RMSprop and gradient descent with impetus. The value of momentum is denoted by beta and is frequently. If you are not absorbed in the mouth behindhand the optimizer, you can just skip.

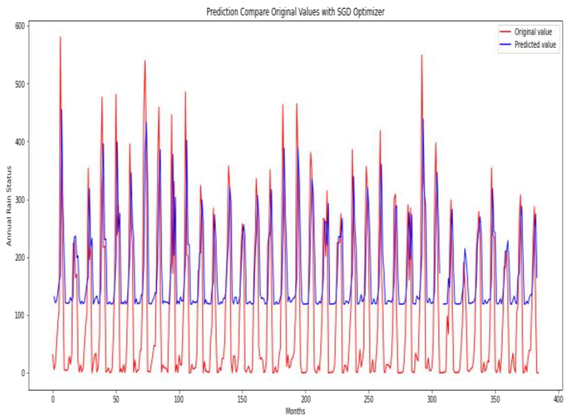

Figure 8: Predicted values compare with original values using SGD optimizer.

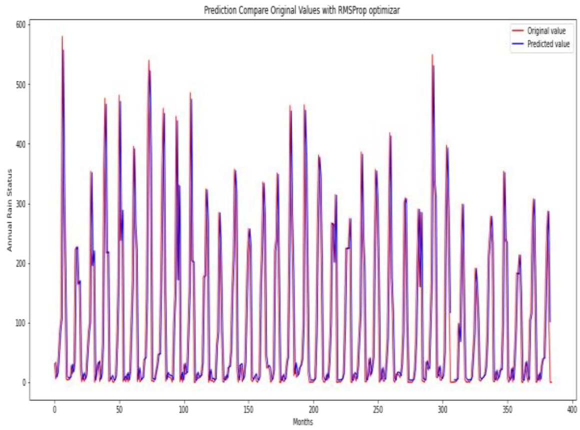

Figure 9: Predicted values compare with original values using RMSprop optimizer.

RMSprop optimizer is a gradient based optimization method used for the exercise neural network. It was planned by father of back-propagation, Geoffrey Hinton. Gradients are actually multifaceted purposes similar neural networks must a propensity to either vanish or detonate as the data spreads finished the function. RMSprop was industrialized as a stochastic method for mini-batch learning. RMSprop contracts with the above issue by using a moving average of squared gradients to normalize the gradient. This regularization balances the step size, decreasing the step for large gradients to avoid explosion, and growing the step for minor gradients to avoid disappearing.

In addition, the accurate result is compared with RSMProp work nearby, our proposed algorithm, as depicted in Figure 10. Further, we try to optimize with ADAM and find the result is much better previous all optimizer. When we merge ADAM optimizer with our propose algorithm, we found predicted values in nearby actual values which are visualized in the graph with the comparison with original values with better accuracy as of previous optimizer.

Adam optimization is the postponement to stochastic gradient first rate and may be utilized in vicinity of classical stochastic gradient descent to replace community masses greater competently. That calculates getting to know charge for every parameter this is proven with the aid of using its builders to paintings properly in exercise and to evaluate favorably in opposition to different adaptive getting to know algorithms.

=

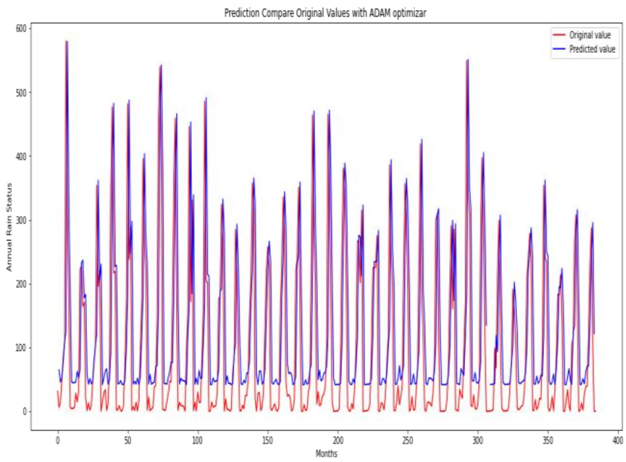

Figure 10: Predicted values compare with original values using ADAM optimizer.

In this figure, researcher, aimed at missing data imputation based on arithmetical prediction, designated in instruction to cover methods applied in the area of missing value imputation. We describe four ways for prediction. One is our propose algorithm and three are with optimization methods. In the research these are compared with original values as depicted in Figure 11. The current work aims to assess for the first time a comparative analysis with SGD optimizer, RSMProp optimizer and ADAM optimizer and also compare with our propose algorithm. There are the graphical representation of the optimizer and propose an algorithm with the original values.

In addition, Figure 12 depicts the accuracy of all scenario which is use in our research, in all optimizer with algorithm is given the best result which display in box plot graph, the ADAM optimizer result is very close to the actual values. This is the conclusion of my result the given graph shows the accuracy of the result we got from all optimizer

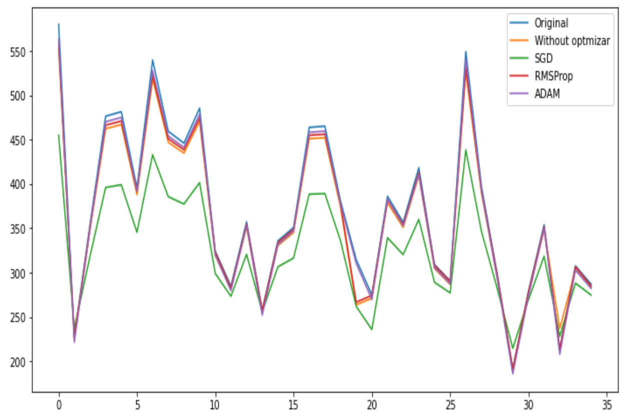

Figure11: Predicted values compare with the original values.

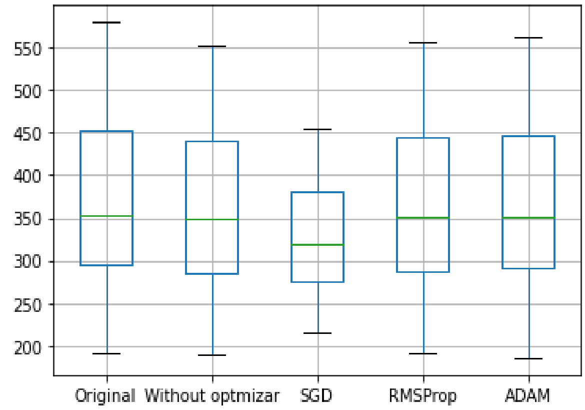

Figure12: Predicted ADAM Optimizer values compare with the original values.

7 Conclusion

In this paper, we are using extended Kalman for missing values imputation it's a new approach for imputation hear we use linear relation for imputation. The assessment of the forecast competence of LSTM based models has the potential to facilitate the further research on the deep learning method projected for rainfall prediction difficulties for ubiquitous computing. We have suggested a loaded bidirectional and unidirectional LSTM architecture for network-wide rainfall forecasting. In addition, the assessment of the forecasting ability of LSTM based models has countless possible to facilitate the strategy of neural network technique for rainfall prediction for ensuring an efficient ubiquitous computing. In addition, the practical rainfall data of Bihar State are tested in this research and the dataset is available on the date. World. Experiments based on one real-world datasets with different missing value specify that the proposed architecture can achieve outstanding imputation and prediction results [25].

Hear is requirement for missing data imputation is the practice of data preparation that has increased popularity last some years for the importance when handling with Time sensor data therefore, research study required some diction approaches for future work. 1.) A comprehensive comparative methodology as imputation techniques with deep learning models. 2.) Growth of a monitoring and warning device to prevent the fault of the sensors. 3.) The implementation of a real-time MVI technique for the holistic predictive framework to assist real-time data-driven decision-making strategies for energy efficient operations of marine machinery.

References:

[1] Wu, C. L., & Chau, K. W. (2013). Prediction of rainfall time series using modular soft computingmethods. Engineering applications of artificial intelligence, 26(3), 997-1007. https://doi.org/10.1016/j.engappai.2012.05.023

[2] Kumar, Dinesh, et al. "A holistic survey on disaster and disruption in optical communication network." *Recent Advances in Electrical & Electronic*

Engineering (Formerly Recent Patents on Electrical & Electronic Engineering) 13.2 (2020): 130-135. https://doi.org/10.2174/2352096512666190215141938

[3] Praveen, Bushra, et al. "Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches." *Scientific reports* 10.1 (2020): 1-21. https://doi.org/10.1038/s41598-020-67228-7

[4] Luengo, J., García, S., & Herrera, F. (2012). On the choice of the best imputation methods for missing values considering three groups of classification methods. Knowledge and information systems, 32(1), 77-108. https://doi.org/10.1007/s10115-011-0424-2

[5] Afrifa-Yamoah, Eben, et al. "Missing data imputation of high-resolution temporal climate time series data." *Meteorological Applications* 27.1 (2020): e1873. https://doi.org/10.1002/met.1873

[6] Ma, J., Cheng, J. C., Ding, Y., Lin, C., Jiang, F., Wang, M., & Zhai, C. (2020). Transfer learning for long-interval consecutive missing values imputation without external features in air pollution time series. Advanced Engineering Informatics, 44, 101092. https://doi.org/10.1016/j.aei.2020.101092

[7] Hassani, Hossein, Mahdi Kalantari, and Zara Ghodsi. "Evaluating the Performance of Multiple Imputation Methods for Handling Missing Values in Time Series Data: A Study Focused on East Africa, Soil-Carbonate-Stable Isotope Data." *Stats* 2.4 (2019): 457-467. https://doi.org/10.3390/stats2040032

[8] Guo, J., Huang, W., & Williams, B. M. (2014). Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. Transportation Research Part C: Emerging Technologies, 43, 50-64. https://doi.org/10.1016/j.trc.2014.02.006

[9] Velasco-Gallego, Christian, and Iraklis Lazakis. "Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study." *Ocean Engineering* 218 (2020): 108261. https://doi.org/10.1016/j.oceaneng.2020.108261

[10] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., &Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), 2222-2232. https://doi.org/10.1109/TNNLS.2016.2582924

[11] Sharma, Ashutosh, et al. "Network modelling and computation of quickest path for service-level agreements using bi-objective optimization." *International Journal of Distributed Sensor Networks* 15.10 (2019): 1550147719881116. https://doi.org/10.1177/1550147719881116

[12] Rhea, S., Wang, E., Wong, E., Atkins, E., & Storer, N. (2017, May). Littletable: A time-series database and its uses. In Proceedings of the 2017 ACM International Conference on Management of Data (pp. 125-138). https://doi.org/10.1145/3035918.3056102

[13] Velasco, Lemuel Clark P., et al. "Week-ahead rainfall forecasting using multilayer perceptron neural network." *Procedia Computer Science* 161 (2019): 386-397. https://doi.org/10.1016/j.procs.2019.11.137

[14] Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y., 2018. Recurrent neural networks for multivariate time series with missing values. Scient. Rep. 8, 6085. https://doi.org/10.1038/s41598-018-24271-9

[15] Chen, X., He, Z., Sun, L., 2019b. A bayesian tensor decomposition approach for spatiotemporal traffic data imputation. Transport. Res. Part C: Emerg. Technol. 98, 73–84. https://doi.org/10.1016/j.trc.2018.11.003

[16] Samal, K. Krishna Rani, et al. "An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach." *Sustainable Cities and Society* 70 (2021): 102923. https://doi.org/10.1016/j.scs.2021.102923

[17] Chen, X., He, Z., Chen, Y., Lu, Y., Wang, J., 2019a. Missing traffic data imputation and pattern discovery with a bayesian augmented tensor factorization model. Transport. Res. Part C: Emerg. Technol. 104, 66–77. https://doi.org/10.1016/j.trc.2019.03.003

[18] Cui, Zhiyong, et al. "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values." *Transportation Research Part C: Emerging Technologies* 118 (2020): 102674. https://doi.org/10.1016/j.trc.2020.102674

[19] Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y., 2018. Recurrent neural networks for multivariate time series with missing values. Scient. Rep. 8, 6085. https://doi.org/10.1038/s41598-018-24271-9

[20] Chaudhry, Aizaz, et al. "A method for improving imputation and prediction accuracy of highly seasonal univariate data with large periods of missingness." *Wireless Communications and Mobile Computing* 2019 (2019). https://doi.org/10.1155/2019/4039758

[21] Chen, Lei, et al. "Comparison of the multiple imputation approaches for imputing rainfall data series and their applications to watershed models." *Journal of Hydrology* 572 (2019): 449-460. https://doi.org/10.1016/j.jhydrol.2019.03.025

[22] Urrea, Claudio, and RaykoAgramonte. "Kalman Filter: Historical Overview and Review of Its Use in Robotics 60 Years after Its Creation." *Journal of Sensors* 2021 (2021). https://doi.org/10.1155/2021/9674015

[23] Liu, Lon-Mu, et al. "Data mining on time series: an illustration using fast-food restaurant franchise data." *Computational Statistics & Data Analysis* 37.4 (2001): 455-476. https://doi.org/10.1016/S0167-9473(01)00014-7

[24] Saputra, M. D., et al. "Handling Missing Values and Unusual Observations in Statistical Downscaling Using Kalman Filter." *Journal of Physics: ConferenceSeries*. Vol. 1863. No. 1. IOP Publishing, 2021. https://doi:10.1088/1742-6596/1863/1/012035.

[25] Jamil, Faisal, and DoHyeun Kim. "Enhanced Kalman filter algorithm using fuzzy inference for improving position estimation in indoor navigation." *Journal of*

Intelligent & Fuzzy Systems Preprint (2021): 1-15. https://orcid.org/0000-0003-1994-6907

[26] Saputra, M. D., et al. "Handling Missing Values and Unusual Observations in Statistical Downscaling Using Kalman Filter." *Journal of Physics: Conference Series*. Vol. 1863. No. 1. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1863/1/012035

[27] Bakibillah, A. S. M., et al. "Robust estimation of traffic density with missing data using an adaptive-R extended Kalman filter." *Applied Mathematics and Computation* 421 (2022): 126915. https://doi.org/10.1016/j.amc.2022.126915

[28] Herrera, Leonardo, et al. "Evolved Extended Kalman Filter for first-order dynamical systems with unknown measurements noise covariance." *Applied Soft Computing* 115 (2022): 108174. https://doi.org/10.1016/j.asoc.2021.108174