Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 28 No. 7s (2025)

Integrating NLP and Indian Traditional Knowledge for Climate Change Mitigation and Adaptation: A Novel Rainfall Prediction Model

[1]Ms. Pradnya Patil, [2]Dr. Sunita Patil

[1]Research Scholar, Department of Computer Engineering,K J Somaiya Institute of Technology, Mumbai, India
pradnya08@somaiya.edu
[2]Director, NMIMS, Deemed to be University, Shirpur Campus, Dhule, India
sunita.patil@nmims.edu

Article History:

Received: 12-12-2024

Revised:28-1-2025

Accepted:9-2-2025

Abstract:

Introduction: The traditional Indian knowledge system encompasses rich intellectual, philosophical, and scientific traditions, mostly documented in Sanskrit, making it less accessible today. This research proposes Sanskrit-to-English Machine Translation to bridge this gap, enabling practical applications. It focuses on translating *Meghamala*, a Sanskrit text on rainfall prediction, to extract key parameters. These insights will be computationally modeled to develop an improved rainfall prediction system. This approach not only enhances weather forecasting but also promotes interest in Sanskrit literature and its scientific relevance in modern society.

Objectives: This research applies NLP-based Machine Translation for accurate Sanskrit-to-English translation, making ancient Indian knowledge accessible. It focuses on developing a rainfall prediction model by extracting meteorological parameters from *Meghamala* and integrating them with modern computational techniques. By bridging traditional wisdom with advanced technology, this approach enhances rainfall forecasting accuracy, benefiting agriculture, water resource management, and climate studies, while also promoting the relevance of Sanskrit literature in scientific research.

Methods: The methodology involves applying NLP-based Machine Translation algorithms to translate *Meghamala* from Sanskrit to English, extracting rainfall-related parameters. These parameters are analyzed and structured for integration into a predictive model. Machine learning techniques are then employed to develop a rainfall prediction system, combining traditional insights with modern meteorological data. The model is validated against historical weather records to assess its accuracy and applicability in real-world forecasting.

Results: The Sanskrit-to-English Machine Translation of '*Meghamala'* can extract rainfall-related parameters, enabling their integration into modern prediction models. This approach enhances rainfall forecasting accuracy, benefiting agriculture and water management. Bridging ancient knowledge with technology, the model demonstrates potential for broader applications in weather prediction and other scientific domains.

Conclusions: Rainfall prediction is crucial for India's economy, which heavily relies on agriculture. Traditional Indian knowledge, documented in Sanskrit, holds valuable scientific insights but remains inaccessible to many. Sanskrit-to-English Machine Translation can bridge this gap, enabling its application in real-world problems like weather prediction, healthcare, and agriculture. The proposed translation of Sanskrit-rich texts like *Meghamala*, which contains extensive rainfall parameters and observations, can aid in developing improved rainfall prediction systems tailored to India’s needs.

Keywords: Climate Change; Indian Knowledge System; Sanskrit; Machine Translation; Rainfall Prediction; NLP

1. Introduction

India, with its vast geography and diverse climatic zones, is highly vulnerable to the impacts of climate change. The country is already witnessing extreme weather events, including erratic rainfall, intense heatwaves, prolonged droughts, and severe flooding. These phenomena pose significant challenges to its agriculture, water resources, biodiversity, and overall socio-economic stability. As a rapidly developing nation with a large population, India faces the dual challenge of maintaining economic growth while addressing the escalating environmental threats posed by climate change. Rainfall prediction plays a pivotal role in agriculture, disaster management, and water conservation, particularly in regions with monsoon-dependent economies such as India. With the increasing unpredictability of weather patterns due to climate change, the need for accurate rainfall prediction has become more urgent. Traditionally, Indian society relied on indigenous knowledge systems, such as the Panchang and the astrological insights found in ancient texts like Varahamihira’s Brihat Samhita, to forecast rainfall.

However, with advances in meteorology and machine learning, the potential for enhancing rainfall prediction has grown significantly. This paper seeks to explore the integration of traditional Indian knowledge systems with modern technology to improve rainfall forecasting accuracy. Such integration holds promise for advancing climate resilience and enhancing agricultural productivity in rural areas. Modelling Rainfall Prediction using Indian Knowledge System and NLP is proposed with amalgamation of study of Indian knowledge System, Sanskrit to English translation and rainfall prediction as research.

1. *Rainfall: An Important Aspect for India***:** Rainfall is vital in developing country, India for economic, social, and environmental benefits, including agriculture, water resources, hydropower generation, food prices, inflation, floods, droughts, biodiversity, and ecosystem health. India's

government and meteorological agencies monitor weather patterns to mitigate extreme events and ensure sustainable water management [1].

2. *Indian Knowledge System:* Language is a communication system used by humans to express thoughts, ideas, and emotions through symbols, sounds, and gestures. India has 121 languages, [2] including Hindi, Marathi, Sanskrit, Tamil, Kannada, Tulu, Punjabi, Nepali, Kashmiri, Bengali, Assamese, and Konkani.

Sanskrit[3] is a language from ancient Indo-Arya with great historical, cultural, and religious significance. With origins that go back more than 3,500 years, Sanskrit is one of the world's oldest languages that has been officially recorded. It is considered the classical language of ancient India and is closely associated with Hinduism, Buddhism, and Jainism. It has influenced numerous languages across the Indian subcontinent. Literary, dramatic, lyrical, devotional, scientific, engineering, and mathematical works are all included in the category of Sanskrit literature. The Sanskrit grammar treatise "Panini Asthadhyayi," penned in the sixth or fifth century BCE by the Indian grammarian Panini, indicates that the Vedic language is highly inflected by nature and portrayed it in the structure of eight stages in a better organised and understandable manner. [4] [5].

The All-India Council for Technical Education (AICTE), a division of the Government of India (GOI), houses the Indian Knowledge System (IKS), an innovation unit within the Ministry of Education (MoE). Its goals are to preserve and spread IKS for future study and social applications, as well as to encourage interdisciplinary research on all facets of IKS. The country's rich legacy and traditional expertise in the fields of the arts and literature, agriculture, basic sciences, engineering and technology, architecture, management, and economics are expected to be aggressively promoted., etc., [6].

Since most of this vast knowledge is based in Sanskrit language it’s not handy for people in today’s era to use it effectively. Thus need a robust solution for this.

3. *Sanskrit to English Machine Translation*: This research proposes Natural Language Processing (NLP) for Sanskrit to English Machine Translation, enabling real-world applications and increasing awareness of Sanskrit literature. This research application proposes to model rainfall predictions using Sanskrit texts, contributing to the preservation, dissemination, and appreciation of the rich cultural and intellectual heritage in Sanskrit literature.

4. *Rainfall Prediction Model:* Rainfall prediction is inherently challenging due to the messy and complex nature of the environment worldwide. While significant progress has been made, there are still uncertainties and limitations in forecasting accuracy, especially for short-term and localized predictions. Researchers and meteorologists continuously work to improve models, data assimilation techniques, and our understanding of atmospheric processes to enhance rainfall predictions.

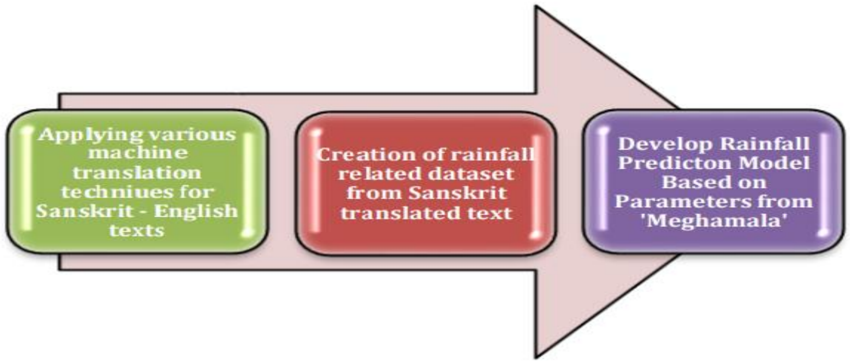

Figure 1: Proposed Methodology

The proposed methodology as shown in Figure 1 as a three-step process and explained below:

1. Apply NLP Machine Translation Algorithm for Sanskrit-English Translation

Natural language processing (NLP) is used for characterizing and investigating human language with computations. It works in combination of Artificial Intelligence, Computer Science and Human Language as shown in Figure 2. Its uses have spread to a variety of industries, including as email spam detection, machine translation, question answering, information extraction, summarization, and the medical field. [7]

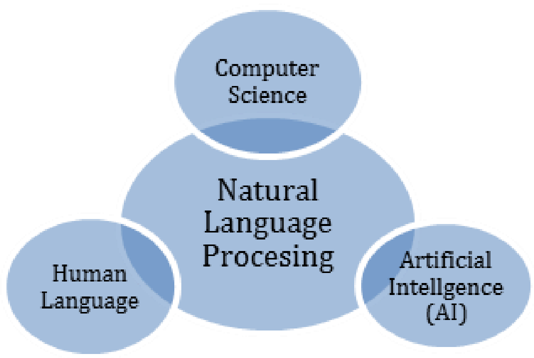

Figure 2: Natural Language Processing: A Part of Human Language, Computer Discipline and AI

It's difficult to make data available to everyone due to language barriers and varying sentence structures. Machine translation, using statistical engines like Google Translate, aims to preserve sentence meaning and grammar. Google recently publicized a fresh approach to machine translation in September 2016 that makes use of deep learning and artificial neural networks. [7]

Machine translation evaluation by human annotators is slow and rigorous, relying on reference translations. Quality estimation eliminates this dependence by estimating translation pairs from source and translated text.

There are four types of Rule-based machine translation; statistical machine translation, example-based machine translation, and neural machine translation are the types of machine translation [8] as depicted in Figure 3:

For Sanskrit-English translation, the choice of machine translation algorithm depends on several factors such as the availability of resources, data, and computational capabilities. Few machine translation algorithms are Transfer Learning with Pretrained Models, Transformer-Based Models, Neural Machine Translation (NMT), Statistical Machine Translation (SMT), and Hybrid Approaches, etc.

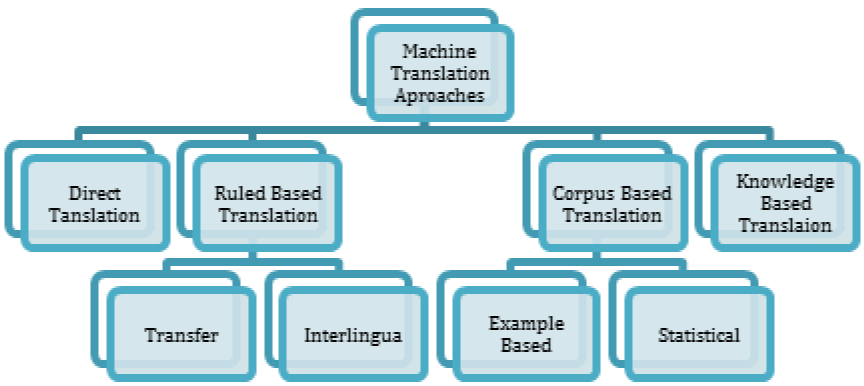

Figure 3: Machine Translation Approaches

Acquire digital Sanskrit text and create dataset.

Digital Sanskrit texts can be accessed through various online resources and libraries, including GRETIL, Sanskrit Documents, Digital Library of India, Sanskrit eBooks, and the Internet Archive. These platforms offer a vast collection of Sanskrit texts, including original works, commentaries, and translations [9] [10] [11] [12] [13]

Creating a dataset for Sanskrit-English texts involves collecting parallel texts in both languages. General process to create a Sanskrit-English dataset are like Identify Reliable Sources, Collect Parallel Texts with corresponding translations in Sanskrit and English, annotate Sanskrit dataset, verify Accuracy and quality of the translations by consulting experts or referring to multiple translations.

2. Develop Rainfall Prediction Model

Rainfall prediction models needs to be designed using various deep learning algorithms based on parameters and rules considered from ‘Meghamala’ Sanskrit text [14]..

This paper is organized under three sections, Section II Literature Review based on Indian Knowledge System, Sanskrit to English Machine Translation and proposed, Rainfall Prediction Model. Section III mentions concepts and design where various kinds of parameters are mentioned which will be used for modelling this rainfall prediction system. Section IV mentions proposed methodology for this research work implementation and at last Section V provides conclusion on this research work.

2. Literature Review

This research is proposed to work under four categories namely Sanskrit as a Profound Language, Study of Sanskrit Literature in Indian Knowledge System, and Natural Language Processing for Sanskrit to English Machine Translation and Scientific Modelling for rainfall prediction.

Conducting a comparative analysis between Sanskrit and the English language reveals fundamental distinctions between these two languages, it functions as an essential first step prior to exploring translation methods. Different elements of the languages are compared, such as number, tense, era, etc.

Era: As a product of the modern era, English is referred to as a modern language. Sanskrit is referred to as a classical language. Its origins date back to the fourth century B.C. [15].

Essence: The development of the language is outlined in Essence. The English language has developed, making it a natural language, but the Sanskrit language was invented by sages like Panini and is therefore artificial or synthetic[15].

Person: A grammatical feature of the language called person allows us to discern "who" is being said. [15]

Sentence formation in Sanskrit and English language varies based on its grammar. For example, a sentence formation in both languages:

Example-1:

Sanskrit: बालः गच्छति।……………………………(i)

English: The child goes………………………(ii)

In this example, the word order in Sanskrit is Subject (बालः), Verb (गच्छति), while in English, it follows the Subject-Verb-Object (The child goes). Sanskrit often allows for a more flexible word order, as the verb can be placed at the end of the sentence.

Example-2:

Sanskrit: तव पुस्तकं पठामि। ……………………. (iii)

English: I read your book. ………………… (iv)

In this example, the word order in Sanskrit is Object (तव) – Verb (पुस्तकं) - Subject (पठामि), while in English, it follows the Subject (I) – Verb (read) - Object (your book). Sanskrit allows for the object to be placed before the verb, whereas English normally follows the Subject – Verb - Object word order.

Different meanings of same Sanskrit sentence:

In NLP, dis-ambiguity refers to the process of resolving ambiguity in language, particularly when a word or phrase has multiple possible meanings or interpretations. Disambiguation can be a challenging task, especially in Sanskrit, as it is an ancient and highly inflected language with rich semantic and contextual variations. [16]

Example-3:

Sanskrit: Ramah Kadalipalam chittva khadantam vanaram pasyai |

English:

1. Rama sees a monkey peeling a banana and eating it.

2. Rama, while peeling a banana, sees a monkey who is eating. [16]

Literature surveyed on Machine Translation in paper [17] uses Semantic Textual Similarity index for Machine Translation Evaluation. They used WordNet semantic feature and worked on CZec - English, Spanish – English, German - English language pairs and used Sagan for MT evaluation.

Promila Bahadur et. al. [15] focuses on how Context Free Grammar and Sanskrit Grammar are comparable to one another, as well as how to build noun words in Sanskrit Grammar using a derivation tree and has created a database for the translation process.

Wei Huangfu, et. al. [18] have created a bilingual corpus that examined Sanskrit analogies and translations of Buddhism into Chinese. With intentions to grow to 100,000 records and 2,000,000 Chinese characters, the corpus currently has 30,000 records and 700,000 characters.

Paper [19] captured functional similarities and encoded long-distance dependencies using a neural technique for translation prediction source dependence-based context representation. Experiments revealed that context-enhanced approaches now in use significantly outperformed baseline systems in large-scale tasks.

Studyin in [20] integrated sentence-level context data into attention-based and Transformer-based NMT using a convolution neural network. Together, the source topic and translation models demonstrated enhanced performance in the method. Experiments revealed notable enhancements over baseline systems.

K. M. Kavitha, et. al. [21] used a tool to suggest word or phrase translation models using a combination of phrase-based and word-based approach. Word to word translation variants for lexicon entries in language pairs were suggested. Their approach involved bilingual stems and suffixes using available translation lexica and suggesting morphological variants for words and phase translations sing learnt bilingual subworld units.

Pawan Goyal et. al. [22] have concentrated on assessing the impact of common parsers using their default configurations. They discovered that in order to efficiently reduce the search space, the L2S model can be expanded by adding linguistic knowledge as constraints in the inference.

Dr. I. Manimozhi et. al. [23] developed a machine translation system for Tulu Script to Kannada Scripts, aiming to improve readability of handwritten Tulu documents. The software effectively recognizes handwritten Tulu characters, achieving an accuracy of 80.5%. The system maps TULU to classical Kannada, preserving valuable data for future generations.

Vishvajit Bakarola and Dr. Jitendra Nasriwala [24] have created a Sanskrit bilingual dataset for use in training a neural machine translation model based on attention mechanisms and language pairs. They have recognised the importance of the attention mechanism in resolving the enduring dependency issue. On Indian languages with little resources, it had encouraging outcomes.

Chaoqun Duan, et. al. [25] proposed a simple method to model future costs in Statistical Machine Translation based on phrases. The authors analyzed on the convergence of Transformer models and compare BLEU score trends between the proposed models and Translation-based models on various validation sets.

Nandini Sethi, Amita Dev and Poonam Bansal [26] designed bilingual automated tool which converts Sanskrit text into English script, allowing users to read without Sanskrit orthography knowledge. Using QWERTY keyboard and Unicode, it supports translation, e-versions, and learning, encouraging human use of original language.

Jivnesh Sandhan, et. al. [27] Our primary focus was Sanskrit word segmentation for machine translation, where we showcased the efficacy of TransLIST as a mutually beneficial solution. TransLIST prioritises possible candidate words, applies an inductive bias to enhance tokenization, and presents a novel path ranking technique to rectify imprecise predictions.

Jivnesh Sandhan, et. al. [28] study focused on the semantic relationships between the components of compound words in order to tackle the Sanskrit Compound Type Identification (SaCTI) task. They put forth a brand-new multi-task learning architecture that combines dependency parsing, morphological labelling, and contextual information. Tests demonstrated the architecture's effectiveness in Marathi and English.

Jivnesh Sandhan, et. al. [29] a neural toolkit for Sanskrit Natural Language Processing (NLP) called SanskritShala1 was presented. The purpose of this toolkit is to facilitate computational linguistic tasks in Sanskrit text, such as compound type identification, dependency parsing, morphological tagging, and word segmentation. With the help of SanskritShala1's user-friendly interactive data annotation tools, annotators may quickly fix system flaws in predictions.

Cuilian Zhang, et. al.'s [30] model for evaluating machine translation faces challenges due to large training data scale and varied quality. Curriculum learning is proposed to rearrange the sequence, using obscurity-quantified frameworks. The proposed methods claim improved segment-level Metrics Shared Tasks compared to baseline methods.

Literature surveyed concludes that the machine translation algorithm needs proper annotated datasets for good results. There are many techniques for machine translation, but their uses differ in language pairs.

Some techniques used by researchers for machine translation need to be checked for various other pair of languages for optimized results. Metrics used for machine translations can be identified and needs to be used for different set of language pairs for their evaluation. Similar to Chinese-Buddhist translation Sanskrit parallels can be constructed which will also help for Pali language works. Irregular word forms and contiguous phrasal forms need to be dealt to further improve the lexicon coverages and can use clear source text topic details be harnessed to enhance the accuracy of predicting target words in Neural Machine Translation (NMT). Neural machine translation using attention based approach can be used for low resource Indic languages. A web-based NLP toolkit can be developed for Sanskrit language featuring post OCR corrections, handling sandhi and meter identification, etc.

3. Concept & Design

There are many Indic Sanskrit texts available which provides information related to Jyotish Shastra, Weather Prediction, Mathematics, etc. Out of these some texts have listing of rainfall prediction system like 'Meghamala'.

The weather prediction in modern systems uses only Atmospheric Indications however, ancient Sanskrit literature mentions four parameters to be considered for weather prediction [31] which are:

a. Parthiva / Bhautik Nimitta - Earthy Indications
b. Antarik Nimitta - Atmospheric Indications
c. Divya Nimitta - Divine Indications
d. Mishra Nimitta - Mixed Indications

This research uses 'Meghamala' Sanskrit text for rainfall predictions. There are various parameters for modelling rainfall prediction however location of predictions based on the observations of the nests and other activities of crow like,

1. Towards east of the tree indicates good rainfall and then ample production of crops.
2. Towards the south-eastern branch indicates possibility of drought.
3. Towards the southern branch indicates absences of rainfall which may lead to quarrel between two states were used in ancient India.

Rainfall Prediction Model:

Modelling rainfall prediction is a complex task that involves understanding the atmospheric processes and interactions that lead to the formation and distribution of precipitation. Various methods and models are used to forecast rainfall, and these can range from simple statistical techniques to sophisticated numerical weather prediction models [32].

1. *Statistical Methods*: These methods are based on historical data and patterns to forecast future rainfall. Techniques like regression analysis, time series analysis, and empirical models are commonly used. While these methods can provide quick and straightforward predictions, they may not capture the underlying physical processes driving rainfall.
2. *Numerical Weather Prediction (NWP) Models*: NWP models are advanced mathematical models that simulate the atmosphere's behaviour based on the laws of physics. These models divide the atmosphere into a three-dimensional grid and solve mathematical equations to predict the evolution of weather variables, including temperature, humidity, pressure, and precipitation. Global models cover large geographic areas, while regional models focus on smaller regions with higher resolution.
3. *Ensemble Prediction Systems:* Ensemble forecasting involves running multiple simulations with slightly varied initial conditions and model physics. These ensembles provide a range of possible outcomes, accounting for uncertainties in the initial data and model representations and help in quantifying forecast uncertainties.
4. *Data Assimilation:* This process involves integrating observed data into numerical models to improve their accuracy. In the context of rainfall prediction, assimilating rainfall observations can enhance the model's performance and reduce errors.
5. *Hydrological Models***:** These models combine rainfall forecasts with land surface and hydrological data to predict how rainfall will impact river flows, water levels, and water resources.
6. *Climate Models:* Climate models, also known as General Circulation Models (GCMs), simulate the long-term behaviour of the atmosphere and oceans. They can be used for seasonal and long-term rainfall predictions and to understand the impacts of climate change on rainfall patterns.

Rainfall prediction is challenging due to atmospheric chaos, with uncertainties and limitations in accuracy. Researchers and meteorologists continuously improve models, data assimilation techniques, and atmospheric processes to improve predictions.

Valmik Nikam et. al.[33] proposed two criteria to evaluate rainfall prediction ability. It tested 11 representative subsets from China Meteorological Administration's open dataset, using three classification algorithms and comparing observing station features and classification algorithms. The prediction accuracy of different subsets is sorted based on station features, examining their influence on accuracy.

Jinghao Niu et. al.’s [34] study from the Indian Meteorological Department (IMD) Pune focuses on rainfall prediction using 7 most relevant attributes. The simulation is prepared using the training data set and it is tested for the accurateness. The model uses data mining techniques to address the issue of compute-intensive rainfall prediction. The Bayesian approach is used to prove the model's effectiveness in rainfall prediction.

E Karthika et. al.'s [35] study found that the method by Holt-Winter's Multiplicative with enhanced smoothing factor offers finer results for rainfall forecasting. The study extracted information from weather historical data from the Indian Meteorological Department, trained a model, and tested its effectiveness using data mining techniques and Bayesian approach.

Yerin Kim et. al.'s [36] For short-term forecasting, the Korea Meteorological Administration (KMA) provided real-time radar observation data. High statistical scores were achieved using a radar rainfall prediction approach based on CGAN for extremely short-range forecasts. The KMA MAPLE system is complemented by this method, which has potential for usage in a number of forecasting applications.

Anik Biswas et. al. [37] developed an algorithm that produced average false alerts and true detection rates of 36.6% and 87% for a 6-hour forecast window based on the GPS PWV data from tropical stations. The suggested threshold values may be applied to long-term forecasting of rainfall with minimum FA rate and better TD rate since they are location-independent.

Jongyun et. al. [38] used Convolution Neural Network (CNN) to forecast rainfall totals using cloud picture data. 85.59% accuracy was attained by the model, with an average mean squared error of 3.05. It is expandable to include rain streaks and more datasets.

Yeji et. al. [39] used Deep learning with advanced enable precipitation nowcasting, with RAIN-F dataset predicting rainfall with 26,280 images and nine atmospheric state variables. It offered superior prediction performance, particularly for heavy rainfall regions, compared to radar data-only experiments.

Mallapa et. al. [40] conducted an analysis of the extended-range prediction for Indian summer monsoon rainfall (ISMR) and evaluated its performance. The forecasting methods employed by the IMD achieved a 64% success rate but faced challenges in accurately accounting for the impacts of ENSO and IOD phenomena.

Prashant et. al.'s [41] study uses IMD rain gauge data to improve rainfall assimilation in the Weather Research and Forecasting (WRF) model using particle filter. The assimilation process involves multiple hypotheses and resampling steps. Results show that particle filter assimilation improves rainfall prediction over CNT runs, and 24-hour forecasts show improved temperature, water vapor mixing ratio, and wind speed profiles.

Umamaheshwari et.al. [42] have used a deep learning model, using a modified LSTM (M-LSTM) approach, to improve prediction accuracy. Their study compared to Naive Bayes, Support Vector Machines, Genetic Algorithms, and Random Forest, revealing its qualitative prediction of rainfall more efficiently.

Manoj et. al. [43] used fitted eight theoretical probability distributions to monthly and annual maximum rainfall data to improve frequency analysis and forecasting. The Gumbel distribution was found to be the best fit for predicting maximum rainfall in Junagadh.

Jitendra et. al. [44] have studied examined historical rainfall and relative humidity trends, finding no significant trends in annual rainfall, Tmax, and Tmin data. There was a positive increasing trend in relative humidity data. Tmax and Tmin showed downward trends, while monthly rainfall increased, with a significant inflection point in a particular year.

Sarabjot et. al. [45] have used monsoon rainfall data in three regions of the state were analysed using non-parametric tests like descriptive statistics, trend analysis, Mann Kendall test, and Sen's slope.

Partha et. al. [46] employed an LSTM neural network to predict monthly rainfall in the Barak River basin of northeastern India, with a lead time of up to 4 months. Our model integrated nine climate variables, encompassing SST, SLP, Nino 3.4 index, ISMR anomalies, and DIPI. The outcomes revealed robust correlations between these predictor sets and the spatial distribution of rainfall, and the model demonstrated impressive performance when validated against real-time observations. This innovative methodology has the potential to significantly enhance the management of water resources in the region.

Ashok et. al. [47] have used methods, including 8-parameter and 10-parameter power regression, and a new statistical ensemble forecasting system, showed potential for improving operational forecasts.

The literature surveyed for rainfall prediction uses different models for different regions based on atmospheric conditions. These predictions are based on historic data from local weather monitoring stations. The rainfall prediction mathematical model can be developed for scalability and different region as universal model. There are some research methods used for these models like deep learning LSTM algorithm, neural networks, Support Vector Machine (SVM), k-nearest neighbour algorithms, etc. for nowcasting, forecasting, etc.

This research proposes to develop the rainfall prediction model based on parameters and observations from various Sanskrit texts.

4. Proposed Methodology

The proposed methodology for developing rainfall prediction model with the use of traditional Indian Knowledge System with Sanskrit book ‘Meghamala’ to model different set of parameters and observations in the atmosphere like bird’s nesting patterns, ant’s movement schedule etc.

The proposed methodology steps are showcased in figure 4.

Rainfall prediction study will be conducted based on IKS. Sanskrit to English translation would be done for very old Sanskrit text 'Meghamala' using various Deep Learning approaches and a new set of parameters and observations will be used from it to develop rainfall prediction model.

In Sanskrit-English machine translation, there is need to work on bilingual dataset. For this dataset creation, applying NLP steps like Lexical Analysis, Syntactic Analysis, Semantic Analysis, Discourse Analysis and Pragmatic Analysis and finally applying various algorithms for translation and evaluation will be studied. For developing rainfall prediction model need of creation of datasets as per parameters from Sanskrit texts and application of deep learning algorithms for better predictions will be identified.

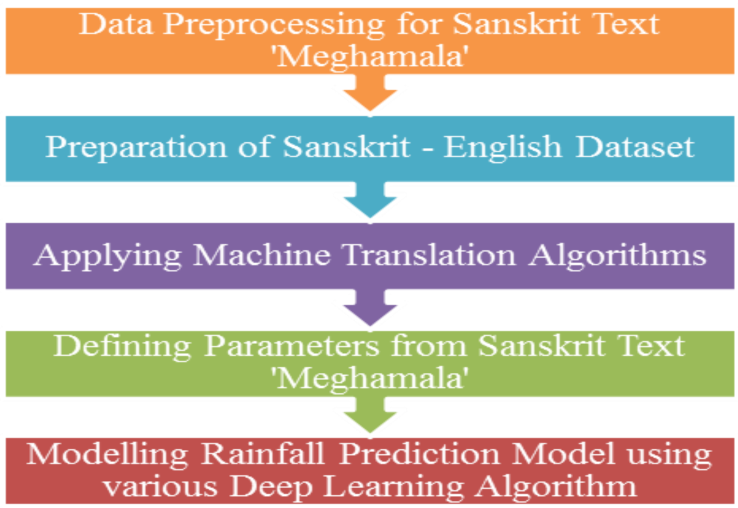

Figure 4: Proposed Methodology Steps to Develop Rainfall Prediction Model

Based on literature survey and the need of Indian economical system, this report proposes the hypothesis to use old Sanskrit texts which have tremendous knowledge about our weather system to create model for rainfall prediction. For this, some very old Sanskrit texts like 'Meghamala' can be used.

5. Conclusion

Rainfall prediction plays a vital role in building economy of the nation. Since, India's economy is majorly dependent on farming and farming results are better in combination with rainfall. Thus, accurate rainfall prediction is a very crucial factor.

The traditional Indian knowledge system encompasses a vast and diverse array of scientific discoveries. Since this vast knowledge is based in Sanskrit language it's not handy for people in today's era to use it effectively. Thus, Sanskrit to English Machine Translation will help many people to get this massive knowledge and apply in real world problems like, Weather Prediction, Healthcare, Agriculture, Education, etc. Thus, the proposed Sanskrit to English machine translation for Sanskrit rich text like ‘Meghamala’ which provides vast parameters and observations for rain fall prediction be modelled to create improved rainfall prediction systems in Indian scenario.

Future Scope:

This research proposal will work on Sanskrit text Meghamala for modelling rainfall prediction. However, the rainfall prediction work can be elaborated by translating other Sanskrit texts from IKS to find more parameters which will help model to improve accurate results of rainfall prediction. The rainfall prediction model will be compared with existing prediction systems from India and abroad for gaining more insights into it.

REFERENCES

[1] India Meteorological Department, “India Meteorological Department: Home.” Accessed: Jul. 27, 2023. [Online]. Available: https://mausam.imd.gov.in/

[2] Wikipedia, “Languages of India.” Accessed: Jul. 09, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Languages_of_India

[3] Wikipedia, “Sanskrit.” Accessed: Jul. 06, 2023. [Online]. Available: https://en.wikipedia.org/wiki/Sanskrit#:~:text=Sanskrit%20(%2F%CB%88s%C3%A6n,in%20the%20late%20Bronze%20Age.

[4] S. C. Kak, “The Paninian approach to natural language processing,” *International Journal of Approximate Reasoning*, vol. 1, no. 1, 1987, doi: 10.1016/0888-613X(87)90007-7.

[5] Book translated by S. C. V. Panini, *The Ashṭādhyāyī of Pāṇini*, 2003rd ed., vol. 1. Motilal Banarsidass Publication, 1891. Accessed: Aug. 23, 2023. [Online]. Available: https://books.google.co.in/books?id=TbzVy6vqyCkC&pg=PP7&source=gbs_selected_pages&cad=2#v=onepage&q&f=false

[6] “Indian Knowledge System.” Accessed: Jul. 06, 2023. [Online]. Available: https://iksindia.org/

[7] D. Khurana, A. Koli, K. Khatter, and S. Singh, “Natural language processing: state of the art, current trends and challenges,” *Multimed Tools Appl*, vol. 82, no. 3, 2023, doi: 10.1007/s11042-022-13428-4.

[8] P. Choudhary and P. Bhattacharyya, “Quality Estimation in MT Evaluation: A Survey.”, Online available: https://www.cfilt.iitb.ac.in/resources/surveys/2023/survey_Param_QE_21June23.pdf

[9] “Göttingen Register of Electronic Texts in Indian Languages”, Accessed: Jul. 09, 2023. [Online]. Available: http://www.sub.uni-goettingen.de/ebene_1/fiindolo/ gretil.htm

[10] "Sanskrit Documents." Accessed: Jul. 09, 2023. [Online]. Available: https://sanskritdocuments.org/

[11] Govt. of India, "Digital Library of India." Accessed: Jul. 09, 2023. [Online]. Available: https://www.dli.gov.in/

[12] "Sanskrit e-Books", Accessed: Jul. 09, 2023. [Online]. Available: http://www.sanskritebooks.org/

[13] "Internet Archive." Accessed: Jul. 09, 2023. [Online]. Available: https://archive.org/

[14] C. T. Sarvanarayan, *Meghamala*, 1993rd ed. GANGANATH JHA KENDRIYA SANSKRIT VIDYAPEETHA, ALLAHABAD, 1993. Accessed: Jul. 25, 2023. [Online]. Available: https://www.exoticindiaart.com/book/details/meghamala-old-and-rare-book-uas167/

[15] P. Bahadur, A. Jain, and D. S. Chauhan, "Architecture of English to Sanskrit machine translation," in *IntelliSys 2015 - Proceedings of 2015 SAI Intelligent Systems Conference*, 2015. doi: 10.1109/IntelliSys.2015.7361204.

[16] S. P. A. K. Malay Maity, "Disambiguation of Instrumental, Dative and Ablative Case suffixes in Sanskrit," in *Proceedings of the Computational Sanskrit & Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference*, Canberra, Australia: Association for Computational Linguistics, Jan. 2023, pp. 71–88. Accessed: Sep. 28, 2023. [Online]. Available: https://aclanthology.org/2023.wsc-csdh.5/

[17] J. Castillo and P. Estrella, "Semantic Textual Similarity for MT evaluation," in *WMT 2012 - 7th Workshop on Statistical Machine Translation, Proceedings of the Workshop*, 2012.

[18] W. Huangfu, Q. Zhu, and B. Qiu, "Construction of a bilingual annotated corpus with Chinese Buddhist translation and their Sanskrit parallels," in *Proceedings of the 2016 International Conference on Asian Language Processing, IALP 2016*, 2017. doi: 10.1109/IALP.2016.7875946.

[19] K. Chen *et al.*, "A Neural Approach to Source Dependence Based Context Model for Statistical Machine Translation," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 26, no. 2, 2018, doi: 10.1109/TASLP.2017.2772846.

[20] K. Chen, R. Wang, M. Utiyama, E. Sumita, and T. Zhao, "Neural Machine Translation With Sentence-Level Topic Context," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 27, no. 12, 2019, doi: 10.1109/TASLP.2019.2937190.

[21] K. M. Kavitha, V. Naik, S. Angadi, S. Satish, and S. Nayak, "Hybrid Approaches for Augmentation of Translation Tables for Indian Languages," in *Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, 2020. doi: 10.1109/ICMLA51294.2020.00157.

[22] A. Krishna, A. Gupta, D. Garasangi, J. Sandhan, P. Satuluri, and P. Goyal, "Neural approaches for data driven dependency parsing in sanskrit," *arXiv*. 2020

[23] I. Manimozhi and M. Challa, "An Efficient Translation of Tulu to Kannada South Indian Scripts using Optical Character Recognition," in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, 2021. doi: 10.1109/ICCMC51019.2021.9418225.

[24] V. Bakarola and J. Nasriwala, "Attention based Neural Machine Translation with Sequence to Sequence Learning on Low Resourced Indic Languages," in *ACCESS 2021 - Proceedings of 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems*, 2021. doi: 10.1109/ACCESS51619.2021.9563317.

[25] C. Duan *et al.*, "Modeling Future Cost for Neural Machine Translation," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 29, 2021, doi: 10.1109/TASLP.2020.3042006.

[26] N. Sethi, A. Dev, and P. Bansal, "A Bilingual Machine Transliteration System for Sanskrit-English Using Rule-Based Approach," 2023. doi: 10.1109/aist55798.2022.10064993.

[27] J. Sandhan, R. Singha, N. Rao, S. Samanta, L. Behera, and P. Goyal, "TransLIST: A Transformer-Based Linguistically Informed Sanskrit Tokenizer," in *Findings of the Association for Computational Linguistics: EMNLP 2022*, 2022.

[28] J. Sandhan *et al.*, "A Novel Multi-Task Learning Approach for Context-Sensitive Compound Type Identification in Sanskrit," in *Proceedings of the 29th International Conference on Computational Linguistics*, 2022.

[29] A. A. L. B. T. S. P. G. Jivnesh Sandhan, "SanskritShala: A Neural Sanskrit NLP Toolkit with Web-Based Interface for Pedagogical and Annotation Purposes," *arXiv:2302.09527v2 [cs.CL] 29 Print Version*, May 2023, Accessed: Jul. 09, 2023. [Online]. Available: https://arxiv.org/pdf/2302.09527.pdf

[30] D. F. W. S. M. I. E. S. K. L. R. Z. and L. S. C. Cuilian Zhang, "Obscurity-Quantified Curriculum Learning for Machine Translation Evaluation," *IEEE/ACM Trans Audio Speech Lang Process*, 2023.

[31] R. V. Sakrikar, "An Indian Perspective Regarding the Science of Weather forecast as Described in Sanskrit Texts," University of Mumbai, Mumbai, 2018. Accessed: Jul. 25, 2023. [Online]. Available: https://shodhganga.inflibnet.ac.in/handle/10603/291678

[32] G. P. Gaikwad and P. V. B. Nikam, "Different Rainfall Prediction Models And General Data Mining Rainfall Prediction Model," *International Journal of Engineering Research & Technology(IJERT)*, vol. 2, no. 7, 2013.

[33] V. B. Nikam and B. B. Meshram, "Modeling rainfall prediction using data mining method: A bayesian approach," in *Proceedings of International Conference on Computational Intelligence, Modelling and Simulation*, 2013. doi: 10.1109/CIMSim.2013.29.

[34] J. Niu and W. Zhang, "Comparative analysis of statistical models in rainfall prediction," in *2015 IEEE International Conference on Information and Automation, ICIA 2015 - In conjunction with 2015 IEEE International Conference on Automation and Logistics*, 2015. doi: 10.1109/ICInfA.2015.7279650.

[35] D. Karthika and K. Karthikeyan, "Analysis of Mathematical Models for Rainfall Prediction Using Seasonal Rainfall Data: A Case Study for Tamil Nadu, India," in *2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2022*, 2022. doi: 10.1109/ICEEICT53079.2022.9768602.

[36] Y. Kim and S. Hong, "Very Short-Term Rainfall Prediction Using Ground Radar Observations and Conditional Generative Adversarial Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3108812.

[37] A. Naha Biswas, Y. H. Lee, and S. Manandhar, "Rainfall Forecasting Using GPS-Derived Atmospheric Gradient and Residual for Tropical Region," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3131217.

[38] J. Byun, C. Jun, J. Kim, J. Cha, and R. Narimani, "Deep Learning-Based Rainfall Prediction Using Cloud Image Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–11, 2023, doi: 10.1109/TGRS.2023.3263872.

[39] Y. Choi, K. Cha, M. Back, H. Choi, and T. Jeon, "RAIN-F: A FUSION DATASET FOR RAINFALL PREDICTION USING CONVOLUTIONAL NEURAL NETWORK," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2021. doi: 10.1109/IGARSS47720.2021.9555094.

[40] M. J. Madolli, S. K. Himanshu, E. R. Patro, and C. De Michele, "Past, Present and Future Perspectives of Seasonal Prediction of Indian Summer Monsoon Rainfall: A Review," *Asia-Pacific Journal of Atmospheric Sciences*, vol. 58, no. 4. 2022. doi: 10.1007/s13143-022-00273-6.

[41] P. Kumar, "Assimilation of the Rain Gauge Measurements Using Particle Filter," *Earth and Space Science*, vol. 7, no. 10, 2020, doi: 10.1029/2020EA001212.

[42] P. Umamaheswari and V. Ramaswamy, "A Novel Modified LSTM Deep Learning Model on Precipitation Analysis for South Indian States," in *Communications in Computer and Information Science*, 2023. doi: 10.1007/978-3-031-27622-4_15.

[43] M. Gundalia, “Best Fitting of Probability Distribution for Monthly and Annual Maximum Rainfall Prediction in Junagadh Region (Gujarat-India),” *MAUSAM*, vol. 74, no. 3, pp. 861–880, Jul. 2023, doi: 10.54302/mausam.v74i3.5898.

[44] J. Rajput, N. Kushwaha, D. Sena, D. Singh, and I. Mani, “Trend assessment of rainfall, temperature and relative humidity using non-parametric tests in the national capital region, Delhi,” *MAUSAM*, vol. 74, no. 3, pp. 593–606, Jul. 2023, doi: 10.54302/mausam.v74i3.4936.

[45] S. Sandhu and P. Kaur, “A case study on the changing pattern of monsoon rainfall duration and its amount during recent five decades in different agroclimatic zones of Punjab state of India,” *MAUSAM*, vol. 74, no. 3, pp. 651–662, Jul. 2023, doi: 10.54302/mausam.v74i3.5331.

[46] P. Sarkar, “Rainfall forecasting in the Barak river basin, India using a LSTM network based on various climate indices,” *MAUSAM*, vol. 74, no. 3, pp. 699–706, Jul. 2023, doi: 10.54302/mausam.v74i3.4933.

[47] A. Kumar, D. S. Pai, J. V. Singh, R. Singh, and D. R. Sikka, “Statistical Models for Long-range Forecasting of Southwest Monsoon Rainfall over India Using Step Wise Regression and Neural Network,” *Atmospheric and Climate Sciences*, vol. 02, no. 03, 2012, doi: 10.4236/acs.2012.23029