

# Fast and Efficient Parallel Alignment Model for Aligning both Long and Short Sentences

Chandramma, Sameena H S, Sandhya Soman, Piyush Kumar Pareek

**Abstract—** Recently, demand for fast and efficient translation system been widely seen. However, translation model are dependent parallel corpora. However, it is challenging to obtain large parallel corpora for resource starved language such as Kannada-Telugu pair. The existing Giza++ based word alignment and Moses phrase based alignment model are efficient for aligning only short sentences. However, for longer sentence the accuracy of model degrades. For performing alignment for longer sentences, neural based alignment has been presented in recent times. However, these models are trained using fixed vector length. Thus, induces memory and training overhead. For overcoming research challenges, this work presents a parallel alignment model using recurrent neural network (RNN). Further, to utilize memory efficiently and minimize training time parallel execution of RNN under GPU is considered. For improving alignment accuracy presented a cost function by combining statistical and neural based alignment method. Experiment are conducted to evaluate the proposed alignment model in terms of accuracy, Word alignment error (WAE), memory utilization, and computation time.

**Index terms—** Neural alignment model, Phrase alignment, statistical alignment, Word alignment model.

## I. INTRODUCTION

Neural Networks (NN) have recently attained wide scope and have been used in applications such as natural language processing (NLP) [1], speech recognition [2], and image processing [3] and attained good result. Since the presentation of these systems and significant outcomes in various applications, numerous scholars in various fields are making utilization of the neural systems as an answer for their issues. MT which is a subcategory of NLP was firstly built utilizing neural systems by [4]. Neural machine alignment and translation model is a new paradigm for alignment and machine translation that has grown superiorly, especially in terms of human evaluation, compared to statistical and rule based alignment and translation model [5], [6] initially developed as a pure sequence to sequence model [7], [8] and enhanced using attention based variant [9]. Neural machine alignment and translation model has now been widely applied for alignment and MT, as well as various other NLP related field such as parsing, summarization, and dialogue.

**Revised Manuscript Received on May 21, 2019**

**Mrs.Chandramma** Department of Computer Science & Engineering at Vivekananda Institute of Technology Bengaluru

**Ms.Sameena** Assistant Professor in Department of Computer Science and Engineering, Global Academy of Technology, Bengaluru

**Ms. Sandhya Soman** Assistant Professor in Kristu Jayanti College, Bengaluru.

**Dr.Piyush Kumar Pareek** Department of Computer Science & Engineering at East West Group of Institutions, Bengaluru.

Number of approaches using neural network has been presented in recent times with good success [11], [12], [13],[14]. However, accuracy is not efficient as compared with phrase based alignment model. Further, neural based alignment and translation model have been used in [15], [16], [17], [18]. In [18] used feed forward neural network for building alignment model. However, [19]] showed RNN attain better performance than feed forward neural network (FFNN). In [20], [21] presented encoder-decoder model using both RNN and convolution neural network (CNN), [22] presented Long Short Term Memory (LSTM) networks for merging statistical alignment model with neural network. In [23], [24], [25] presented an alignment model to address memory issues for performing aligning longer sentences. Further, [26], [27] presented bi-directional alignment and translation model. However, these model incurs computation overhead for performing decoding operation (i.e., due to poor convergence rate when compared with statistical alignment model).

The methodology presented so far for neural based alignment and translation model are generally belongs to encoder-decoder family and encodes all necessary information of source sentence into fixed-length vector from which the decoder performs alignment and translation. This may affects the alignment of neural system for longer sentences, especially those sentence that longer than the sentences in the training dataset. To address, this paper present a novel alignment model that automatically soft search parts of source sentence that are important to predicting target phrase/word, without necessity to form these part as a hard segment explicitly. The most important feature of our approach is it does not encode a whole sentence into fixed-length vector. Instead, it encode sentence into set of vectors and select subset of these vector in adaptive manner in decoding process. This enables neural system to eliminate having all information of source sentence, irrespective of its length, into fixed-length vector. Further, memory overhead of neural based model for aligning longer sentence is addressed using parallel recurrent neural network.

For machine translation, neural system [5] has been widely used across various language pairs. In this work, we present a neural network alignment model for Dravidian language (Kannada-Telugu pair). According to our knowledge, this is the first neural network based alignment model for Dravidian language. We improve the base phrase based

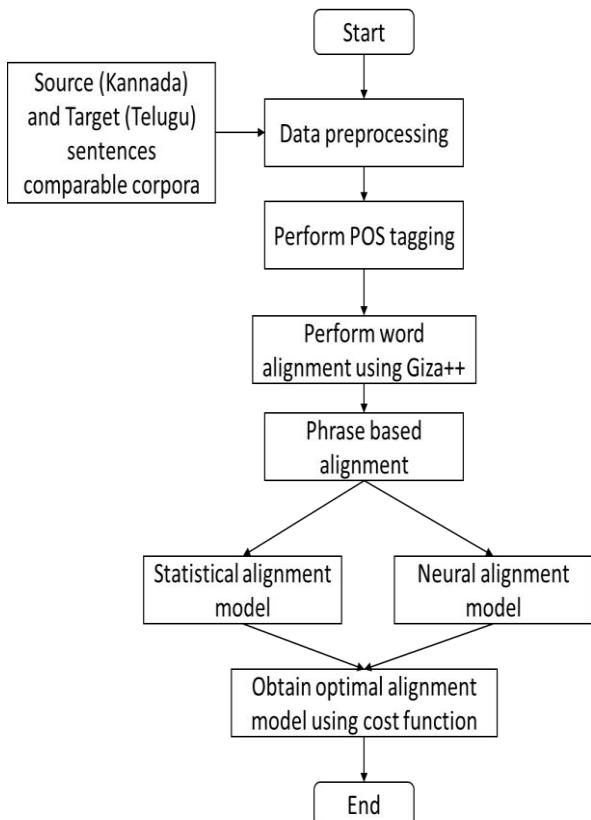


model [5] developed by Google with new feature obtained from statistical phrase based alignment model [10]. The new model is composed of new parameter as a cost function that measure variance among alignment obtained from our phrase based alignment model and neural alignment model. Then this cost function is used to enhance convergence time and accuracy of neural based alignment model. Along with minimize memory overhead (MO) and achieve superior alignment accuracy for longer sentence with minimal computation time (CT).

### Research Contribution are as follows:

- Here we present a novel phrase based alignment model for Kannada-Telugu pair.
- Presenting a novel cost function to obtain better alignment performance among statistical and neural network based alignment model.
- Our model achieves higher alignment accuracy, better memory utilization and with minimal computation overhead.

The rest of the paper is organized as follows. In section II the proposed alignment architecture is presented. In section III the proposed model is presented. In penultimate section experimental study is carried out. The conclusion and future work is described in last section.



**Fig. 1.** Proposed alignment model architecture

## II. PROPOSED ALIGNMENT MODEL ARCHITECTURE

This section present the architecture design of proposed alignment model which is shown in fig. 1. In data

preprocessing process the comparable corpora of Kannada-Telugu pair is collected from Emille corpus. Further, the corpora collected are tagged using method presented in [35]. Then, word alignment is done on tagged data using Giza++ alignment tool. Then phrase alignment is done using statistical alignment model presented in [10]. Further, phrase alignment is done using neural alignment model presented in this work. Then, to obtain optimal alignment this work further presented a novel cost function by combining both statistical and neural alignment model to obtain better alignment considering both long and short sentence.

## III. FAST AND EFFICIENT PARALLEL ALIGNMENT MODEL FOR BOTH LONG AND SHORT SENTENCES

This paper present a multilayer recurrent neural network based alignment model for Dravidian language pairs. And also present a new cost function model using statistical alignment model [10] for achieving better convergence and accuracy. The multilayer recurrent neural network is composed of encoder and decoder, an encoder reads input sentence which is a sequence of vectors  $a = (a_1, \dots, a_{U_a})$ , into a vector  $d$ . Here we consider variable length sentences into variable length vectors rather than fixed-length vectors for better alignment performance using RNN such that

$$i_u = g(a_u, i_{u-1}) \quad (1)$$

and

$$d = r([i_1, \dots, i_{U_a}]), \quad (2)$$

where  $i_1 \in S^q$  is a hidden state at instance  $u$ ,  $g$  and  $r$  are some nonlinear functions, and  $d$  is a vector obtained from set of the hidden states.

The decoder is generally trained to compute the next word  $b_u$  for previously computed words  $\{b_1, \dots, b_{u-1}\}$  and for a given context vectors  $d$ . The role of decoder is to outline a probability over the alignment  $b$  by changing the joint probability into the ordered conditional as follows

$$\mathcal{P}(b) = \prod_{u=1}^U \mathcal{P}(b_u | \{b_1, \dots, b_{u-1}\}, d), \quad (3)$$

where  $y = (b_1, \dots, b_U)$ . With a multilayered RNN, each conditional probability is defined as follows

$$\mathcal{P}(b_u | \{b_1, \dots, b_{u-1}\}, d) = h(b_{u-1}, t_u, d), \quad (4)$$

where  $h$  is potentially a multilayered nonlinear function that obtains the probability of  $b_u$ , and  $t_u$  is the hidden state of multilayered recurrent neural network.

Now we present a novel parallel alignment model that emulates searching through source data during decoding alignment process. Here we describe each conditional probability in Eq. (3) as follows



$$\mathcal{P}(b_m | \{b_1, \dots, b_{m-1}\}, x) = h(b_{m-1}, t_m, d_m). \quad (5)$$

where  $t_m$  is a hidden state of RNN for instance  $m$ , which is computed as follows

$$t_m = g(b_{m-1}, b_{m-1}, d_m) \quad (6)$$

It can be seen from Eq. (3) unlike traditional model here the probability is conditional on distinctive context vector  $d_m$  for each word  $b_m$ . The context vector  $d_m$  depends irrespective of the context (or sequence of annotations)  $(i_1, \dots, i_{U_a})$  to which encoder maps the input data. Each  $i_m$  is composed of information of entire input sequence with main optimization around  $m^{\text{th}}$  word of the input corpus data. Then the context vector  $d_m$  is then computed as a weighted sum of these  $i_m$ .

Further, this work present a novel cost function for obtaining optimal alignment considering diverse data of both short and long sentences. For achieving, we presented a cost function combining statistical alignment model and neural network based alignment model. For statistical alignment model, we use model presented in [10] to align the source to target sentences. This model uses phrase based alignment and the alignment can be described as the following matrix

$$\mathcal{A}^{T \times S} = \begin{cases} \mathcal{A}[m, n] = 1 & \text{If the } m^{\text{th}} \text{ target phrase is aligned to the } n^{\text{th}} \\ & \text{source phrase} \\ \mathcal{A}[m, n] = 0 & \text{Otherwise} \end{cases} \quad (7)$$

Here,  $T$  and  $S$  is the size of the target and source sentence, and  $\mathcal{A}$  is the alignment matrix obtained using [10]. In neural based alignment, soft alignment is used. So we have one more matrix, namely, Neural based alignment matrix which is the outcome of proposed neural based alignment model.

Now we define a cost function to alignment model which is composed of difference among both matrix. This steps is considered here to attain better neural based alignment model i.e., the neural based alignment model will converge faster and at same instance it finds the alignment model which is suitable for corpora. Since the alignment model [10] has a better alignment model than neural based alignment model which is in the process of training and is not convergent. The cost function for attaining better alignment and translation is computed as follows

$$\frac{\alpha}{2(|T| + |S|)} * W \quad (8)$$

where

$$W = F(|\mathcal{A}^{T \times S} - f^{T \times S}|) \quad (9)$$

In the above Eq. (8) and (9),  $f$  is the neural based alignment matrix and  $\mathcal{A}$  is the statistical alignment matrix,  $W$  represent absolute value, and function  $F$  is summation of entire matrix element. The term that is used just before the functions is for normalizing the summation process. And  $\alpha$  is weight describing how significant this term over default cost function. Higher the weight of  $\alpha$ , more significant the alignment transformation is. In this paper, the weight is set to 0.2 because the cost function is composed of difference among model alignment and target alignment and this

difference is more significant than the alignment difference. The cost function is used for learning rate decay influence. In neural based alignment at each step the model is expected to be decrease the cost function. If after set of iterations, the cost function is not changed or decreased, the learning rate will be changed. Adding a new term to cost function will aid in improving the alignment of neural based alignment model and is expected to have alignment closer to statistical alignment model. In next section, the performance of proposed alignment is evaluated over stat-of-art technique.

#### IV. EXPERIMENT AND RESULT ANALYSIS

This section present performance evaluation of proposed parallel alignment model over state of alignment model. The performance of model is evaluated in terms of decoding time. Here we train our model considering sentence up to 50 words. The encoder and decoder is composed of 1000 hidden units each. The encoder of alignment searching model is composed of forward and backward recurrent neural network each having 1000 hidden units. Here we consider multilayer network maxout [29] hidden layer to evaluate the conditional probability of each target word [30]. We consider adadelta [31] along with mini-batch stochastic gradient descent algorithm to train our model. Each stochastic gradient descent update direction is evaluated using a mini-batch of 100 sentences. We trained our model for approximately 4 days and we use beam search to identify alignment that maximizes the conditional probability.

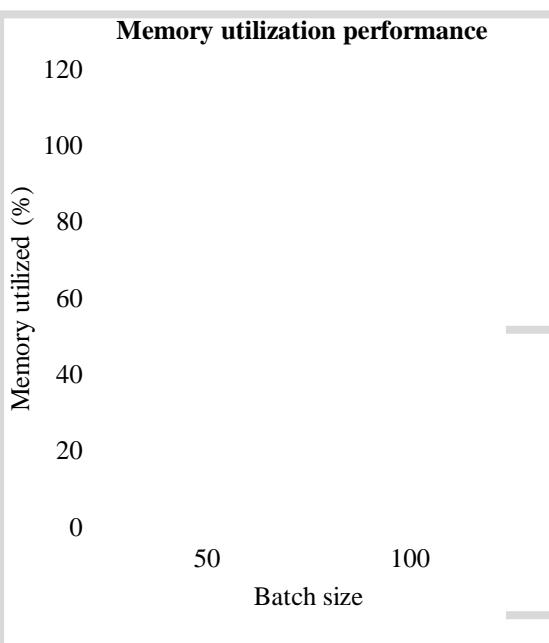
For performance evaluation for Dravidian language, the dataset are obtained from Emille corpus [33]. The training data is composed of 4.2 million sentence pairs. We consider sentence with maximum size of 50 words and eliminate duplicate sentences pairs. Post filtering we have parallel sentence of 4 million sentence pair. For testing alignment, the sentences is composed of 80000 words, out of which first 64000 were taken as training data and the remaining 16000 words were taken as test data. The test data contained 40% of words that were not part of training data. The system environment to run these corpus are run on Intel I-7, 3.5GHz, quad core class processor, 16GB memory, trained on one CUDA enabled GPU NVIDIA GeForce 820M, CUDA version 8.0, and cuDNN version 5005. Experiment outcome obtained from table I shows, the alignment model [5] attain a decoding time of 0.2118 for English to France alignment. However, the proposed alignment model attain a decoding time of 1.6980 for Kannada-Telugu language pair. This work is the first neural network based alignment model for Kannada-Telugu language pair. Further, the significance of proposed alignment model is it can align sentence composed of word greater than 50 word with accuracy of 91.95% and word alignment error (WAE) of 9.78%. Further, memory utilization performance is evaluated by varying the batch size from 50 and 100 and result obtained is graphically shown in Fig. 2. From result obtained it can be seen, an



average memory utilization and wastages of 95.84% and 4.16% is attained by proposed parallel alignment model under GPU environment, respectively. The overall result attained shows efficiency of proposed alignment model over existing alignment model [5], [9], [34].

*Table 1* Performance evaluation

	Existing model [5]	Proposed model
Number of word per sentence	-	16
Batch size sentence	50	100
Accuracy	-	91.95%
WAE	-	9.78%
CPU decoding time	0.2118	1.6980
Memory utilization	-	95.84%
Memory wastage	-	4.16%
Language pair considered	English to France	Kannada-Telugu



**Fig. 2. Memory utilization performance evaluation considering varied batch size**

## V. CONCLUSION

This work present an efficient alignment model for various Kannada-Telugu language pair. Our model can be used to train various language pair in a scalable manner by fully utilizing the GPU architecture. Our model attain good training efficiency due to better memory management by using GPU. As a result improving the decoding time. Our model attains good alignment accuracy and word alignment error performance compared to state-of-art techniques. The

improvement of accuracy is due to new cost function defined for performing alignment. The cost function aid in improving convergence rate of neural network based alignment. An average performance alignment accuracy of 91.95%, word alignment error of 9.78%, and memory wastage of 4.16% is achieved by proposed alignment model. Future work would further consider improving the accuracy of alignment by optimizing training parameters and training consider large training sets. Along with consider performing experiment on different language pair.

## REFERENCES

1. Bollen, H. Mao, X. Zeng. "Twitter mood predicts the stock market." Journal of Computational Science. Vol. 2(1), pp. 1-8, March 2011.
2. G. Hinton, L. Deng , D. Yu, GE. Dahl, AR. Mohamed, N. Jaitly, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". IEEE Signal Processing Magazine, vol. 29(6), pp. 82-97, Nov. 2012.
3. D. Ciregan, U. Meier, I. Schmidhuber, "Multi-column deep neural networks for image classification." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference, pp. 3642-3649, Jun. 2012.
4. MA. Castano, F. Casacuberta, E. Vidal, "Machine translation using neural networks and finite-state models," Theoretical and Methodological Issues in Machine Translation, pp. 160-167, Jul. 1997.
5. Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation, arXiv:1609.08144, 2016.
6. Josep Crego, Jungi Kim, and Jean Senellart. Systran's pure neural machine translation system, arXiv:1602.06023, 2016.
7. Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to Sequence Learning with Neural Networks. In NIPS. page 9. arxiv:1409.3215, 2014.
8. Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In EMNLP, 2014.
9. Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective Approaches to Attention based Neural Machine Translation. In EMNLP, 2015.
10. Chandramma, P. Kumar Pareek, K. Swathi and P. Shettepanavar, "An efficient machine translation model for Dravidian language," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, pp. 2101-2105, 2017.
11. Koehn, P., Och, F. J., and Marcu, D. Statistical phrase-based translation. In Conference of the North American Chapter of the Association for Computational Linguistics, 2003.
12. Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R. M., and Makhoul, J. Fast and robust neural network joint models for statistical machine translation. In ACL (1), Citeseer, pp. 1370–1380, 2014.
13. Kalchbrenner, N., and Blunsom, P. Recurrent continuous translation models. In Conference on Empirical Methods in Natural Language Processing 2013.
14. Cho, K., van Merriënboer, B., Gülcöhre, Ç., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Conference on Empirical Methods in Natural Language Processing 2014.
15. Y. Bengio, "A neural probabilistic language model," The Journal of Machine Learning Research, Vol. 3, pp. 1137-1155, 2003.
16. I. Sutskever, O. Vinyals, QV. Le, "Sequence to sequence learning with neural networks," In Advances in neural information processing systems, pp. 3104-3112, 2014.
17. T. Mikolov, "Statistical language models based on neural networks," Presentation at Google, Mountain View, April 2012.

18. LH. Son, A. Allauzen, and Fr. Yvon, "Continuous space translation models with neural networks." Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: Human language technologies. Association for Computational Linguistics, pp. 39-48, Jun 2012.
19. M. Sundermeyer, I. Oparin, JL. Gauvain, B. Freiberg, R. Schlüter, and H. Ney, "Comparison of feed forward and recurrent neural network language models." Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. IEEE, pp. 8430-8434, May 2013
20. N. Kalchbrenner, and B. Phil, "Recurrent Continuous Translation Models," EMNLP, Vol. 3, p. 413, 2013.
21. M. Auli, M. Galley, C. Quirk, and G. Zweig, "Joint Language and Translation Modeling with Recurrent Neural Networks," EMNLP. Vol. 3, pp. 1044-54, 2013.
22. K. Cho, B. Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," In Proceedings of the Empirical Methods in Natural Language Processing, pp. 1724-34, Jun. 2014.
23. L. Pouget-Abadie, D. Bahdanau, B. van Merriënboer, K. Cho, and Y. Bengio, "Overcoming the curse of sentence length for neural machine translation using automatic segmentation," arXiv preprint arXiv:1409.1257. Sep. 2014.
24. H. Schwenk, "Continuous Space Translation Models for Phrase-Based Statistical Machine Translation," COLING (posters), pp. 1071-1080, Dec. 2012.
25. Devlin, R. Zbib, Z. Huang, T. Lamar, RM. Schwartz, and I. Makhoul, "Fast and Robust Neural Network Joint Models for Statistical Machine Translation," InACL Vol. I , pp. 1370-1380, Jun. 2014.
26. M. Schuster and KK. Paliwal, "Bidirectional recurrent neural networks," IEEE Transactions on Signal Processing, Vol. 45, pp. 2673-81 , Nov. 1997.
27. M. Sundermeyer, T. Alkhouri, I. Wuebker, and H. Ney, "Translation Modeling with Bidirectional Recurrent Neural Networks," InEMNLP 2014, pp. 14-25, Oct. 2014.
28. W. Chen, E. Matusov, S. Khadivi, JT. Peter, "Guided Alignment Training for Topic-Aware Neural Machine Translation," arXiv preprint arXiv:1607.01628, 2016.
29. Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., and Bengio, Y. Maxout networks. In International Conference on Machine Learning, pages 1319– 1327, 2013.
30. Pascanu, R., Gulcehre, C., Cho, K., and Bengio, Y. How to construct deep recurrent neural networks. In Second International Conference on Learning Representations, 2014.
31. Zeiler, M. D. ADADELTA: An adaptive learning rate method. arXiv:1212.5701, 2012.
32. Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi. Findings of the 2015 Workshop on Statistical Machine Translation. In Tenth Workshop on Statistical Machine Translation, September 2015.
33. EmilleCorpus. <http://www.lancaster.ac.uk/fass/projects/corpus/emille/>.
34. Dakwale, Praveen & Monz, Christof. Convolutional over Recurrent Encoder for Neural Machine Translation. The Prague Bulletin of Mathematical Linguistics. 108. 10.1515/pralin-2017-0007, 2017.
35. Chandramma and Piyush Kumar Pareek. Fast and Accurate Parts of Speech Tagging for Kannada-Telugu Pair. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 10 (2018) pp. 7857-7867, 2018.

Technology Bengaluru and pursuing her Ph.D in Computer science & Engineering, VTU Belagavi. Her Interested areas of Research are Natural Language Processing, Data mining & Machine Learning.



Ms. Sameena H S received the B.E degree in Information Science and Engineering and M.Tech degree in Computer Science and Engineering from VTU, Belgaum in 2008 and 2013 respectively, and currently working toward Ph.D. degree from VTU, Belgaum. Currently working as Assistant Professor in Department of Computer Science and Engineering, Global Academy of Technology, Bengaluru and total teaching experience is 10 Years. Areas of interest are Software Engineering, IoT, and Image Processing.



Ms. Sandhya Soman received Master's Degree in Computer Applications from Guru Gobind Singh Indraprastha University, New Delhi and is currently pursuing her PhD in Computer Science. She is currently working as an Assistant Professor in Kristu Jayanti College, Bengaluru. Her research areas include Agile and lean Software Development.



Dr. Piyush Kumar Pareek has received Ph.D (2016) in Computer Science Engineering from Jain University, Bengaluru, M. Tech (2012) from Dayanand Sagar College of Engineering, Bengaluru & B.E (2010) from Basaveshwar Engineering College, Bagalkote, Author has over 8+ Years of Academic Experience, He is author or co-author of over 53 international journal papers or conference proceedings, He has Four Books published in Lambert Publications, Germany. He is BOS Member – Alumni – ISE – of Basaveshwar Engineering College, Bagalkote, He is PhD Alumni Convener of Jain University, Bengaluru. He is Recognized Research Supervisor in Computer Science & Engineering (VTU0817401), Visvesvaraya Technological University, Belagavi. He is currently serving in department of Computer Science & Engineering at East West Group of Institutions, Bengaluru. His Interested areas of Research are Software Engineering, Computer Networks & Machine Learning. He is Editorial Board Member & Reviewer of International Journal of Advanced Research in Computer Science & Technology (IJARCST), Brittany, FRANCE. Reviewer of International journal of Engineering Research and Technology ESRSA Publications (IJERT), INDIA, International journal of Computer Science Research and Technology (IJCST), INDIA, Inter Science Research Network (IRNET), Bhubaneswar, INDIA, International Journal of Computer Science and Informatics (IJCSI), Bhubaneswar, INDIA, International Journal of Engineering, Sciences and Management (IJESM), U.P, INDIA, International Journal of Scientific & Technology Research (IJSTR), Paris, FRANCE, Journal of Theoretical and Applied Computer Science (JTACS), Szczecin, Poland, International Journal of Advanced Research in Computer Science & Technology (IJARCST), INDIA, Journal of Harmonized Research in Engineering, INDIA, IASTER's International Research Journals, INDIA, IJMAS Journals, Bhopal, INDIA, RR Journals, INDIA, Reviewer of IRD INDIA & STEM Journals, INDIA.

## AUTHORS PROFILE



Mrs. Chandramma has received Master's degree in Computer science & Engineering from VTU, Belagavi in 2006, B.E in CSE from Bangalore University in 1999. Author has over 18 years of Academic experience. She has published one book in Lambert Publications, Germany and published several papers in International Journals and also presented & published several papers in International or national conferences. She is currently serving in Department of Computer Science & Engineering at Vivekananda Institute of

