Forecasting The Adoption Rate Of The E Learning Using Multilayer Recurrent Neural Network With Long Short Term Memory On Analysis Of User Sentiment

S. Sirajudeen[1*], B. Vivekanandam[2], Haleema Pk[3]

[1*]Research scholar, Faculty of Computer Science and Multimedia, Lincoln University College, Malaysia,
Email: sirajudeenhod@gmail.com
[2]Deputy Dean, Faculty of Computer Science and Multimedia, Lincoln University College, Malaysia,
Email: vivekanandam@lincoln.edu.my
[3]Senior Lecturer, School of Computing and Engineering, University of West London, RAK Branch, Campus, UAE,
Email: contact@drhaleema.in

Abstract
E learning has become a necessary option to the education system to the entire world due to occurrence of the covid -19 pandemic. Especially implementation of the lockdown, educational institution and students adapted to e-learning system. In order to evaluate the opinion of the students to e- learning system with respect to various challenges and advantages, sentiment analysis has been employed is used to gain valuable insight. The social networks are widely distributed to gather and share the user perspective. This textural information is highly sourced with the data providing the feelings of the students with the statements that expresses agreement or disagreement in the comment sections to reveal the negative or positive feelings of the students towards the learning for performing the sentiment analysis and opinion mining. With current technology advancement in the information technology, it is necessary to forecast the adaption rate of the e-learning system in future. Therefore, a new optimized Multilayer recurrent neural network with long short term memory architecture is proposed in this work to provide valuable forecasting and suggestion on the user intension. Model is adaptable to temporal and time varying data along its existence of long term dependences between the users. Initially data preprocessing is carried out user profiling. Preprocessed data is employed for Latent Dirichlet Allocation feature to extract the opinion on the latent user intention process on the parsed data and store long dependency of the data in latent representations on the LTSM Model. Latent user intention is projected to the long short term memory for efficient data organizing on indexing as network. Organized data is employed to Multilayer Recurrent Network composed of various layer is considered as learning representative which provides valuable forecasting as suggestion and recommendation on various aspects on user adoption. We evaluate the performance of the proposed deep learning approach on a twitter dataset which considered of 38602 tweets during covid -19 pandemic. It has been proved to be outperforming against state-of-the-art methods against precision, recall and F Measure respectively.

Keywords: E-Learning, Adaptation Rate, Long short term memory, Recurrent Neural Network, Latent Dirichlet Allocation, Sentiment Analysis

1. Introduction

Sentiment analysis is natural processing technique which identifies and extracts the subjective information of the data. It is employed to various domains to obtain the user perception on the particular context. Especially sentiment analysis helps to forecast and recommend the intention of the user in the social media[1]. In specific, sentiment analysis finds its importance in the forecasting the solutions in the e learning systems. E learning has become a necessary option to the education system to the entire world due to occurrence of the covid -19 pandemic [2]. Especially implementation of the lockdown, educational institution and students adapted to e-learning system. In order to evaluate the opinion of the students to e- learning system with respect to various challenges and advantages, sentiment analysis has been employed is used to gain valuable insight on basis of learners interaction.

The social networks are widely distributed to gather and share the user perspective on the multiple emotional dispositions within the context[3]. This textural information is highly sourced with the data providing the feelings of the students with the statements that expresses agreement or disagreement in the comment sections to reveal the negative or positive feelings of the students towards the learning for performing the sentiment analysis and opinion mining[4]. With current technology advancement in the information technology, it is necessary to forecast the adaption rate of the e-learning system in future. Therefore, a new optimized Multilayer recurrent neural network with long short term memory architecture is proposed in this work to provide valuable forecasting and suggestion on the user intension.

Model is adaptable to temporal and time varying data along its existence of long term dependences between the users. Initially data preprocessing using stop word removal and stemming along work parsing. Preprocessed data is employed for Latent Dirichlet Allocation[5] feature to extract the opinion on the latent user intention process on the parsed data and store long dependency of the data in latent representations on the LTSM Model[6]. Latent user intention is projected to the long short term memory for efficient data organizing on indexing as network. Organized data is employed to Recurrent Network composed of various layer is considered as learning representative which provides valuable forecasting as suggestion and recommendation on various aspects on user adoption.

The Remaining paper is categorized as follows, related work are described in section 2, the architecture of the proposed long short term recurrent neural network is described in section 3 and experimental results and effectiveness of the proposed system is demonstrated in section 4 using real time dataset along performance comparison against state of arts approaches on various metric has been explained. Finally paper will be concluded in section 5.

2. Related works

In this section, sentiment analysis technique using machine learning approaches has been examined in detail on basis of techniques for latent feature extraction and opinion classification along forecasting. Each of those machine learning models which produce good performance with respect to accuracy and effectiveness has been represented in detail and few techniques which performs nearly similar to the proposed architecture is analysed as follows

2.1. Sentiment Analysis for E-Commerce Product Reviews

In this model, sentiment lexicon is employed to extract and enhance the sentiment features on the various contexts in the review[7]. Extracted lexicon is classified on basis of polarity using machine learning architectures. Feature vector contain the Sentiment feature as word embedding model and those feature is classified on basis of the Feature weighting using Support Vector Machine or KNN classifier.

2.2. Semi-supervised multitask learning

In this method, aspect based sentiment analysis is employed to extract the work describing the aspect of the entity. Those extracted work will be analysed with sentiments on processing it in vector form. Vector is analysed on dependencies structures and exploit the indirect relationship on the opinion words. Opinion words are allocated with sentiments using decision tree classifier such as Naive bayes or c4.5 algorithm[8].

3. Proposed model

This section provides a detailed design specification of the proposed technique titled multilayer Recurrent Neural Network[9] with long short term recurrent neural network as deep learning architecture employed for forecasting the adaption rate of the learners to the e learning system on inclusion of parametric tuning of the layers using error function to obtain the effective prediction of the recommendation.

3.1. Problem statement

Forecasting has become essential to determine the opinion of the learner as their response of the learning. Opinion data has been limited only specified time frame. The online learning community is trying to reproduce that generic model for any kind of time frame in order to improve the efficacy of forecasting and recommendation. Furthermore, novel solution required to forecast the adaption rate of the learner in e learning system for detectable changes in the online learning.

3.2. Data Preprocessing –User profiling

E learning system which consists of instructor pool, student pool and instructor allocate the materials to the students[10]. In order to extract the opinion of the student, student profiling has to be obtained as follows.

$U = \{u_1, u2, u3....u_n\}$ Eq.1

- **User Profiling**

The User profiling keeps track of opinion and intention for each student in the e learning system including various contexts with respect to time in doing tasks in each category. These intentions are dynamically maintained, and can be used to guide the sentiment analysis model[11] to determine the sentiment and their adaptation rate of the student for future learning. Moreover, the user profiling contains the information of the user with specified constraints.

$D= \{d_1, d_2\ d_3....d_n\}$...Eq.2

- **Adaption Estimation**

In order to avoid the intrinsic error rate, weighted voting based approach has been implemented on the opinion to user experience to the system on intention aggregation of the user on basis of learning context. Especially, it is to determine large no of classes to the user opinion with respect to sentiment to the various context on basis of the user behaviour and user experience[12] to the context.

Furthermore latent discriminant analysis has been employed to determine the latent factor of developer for the user. User profile has been illustrated in the matrix form as complete user perception matrix. The Complete user perception matrix composed of column mentioning the data point of the user on characteristics and row mentioning the user behaviour characteristics or experience. It further employed to minimize over fitting issue by computing the Final projection matrix [13].

Aggregated User behavior Vector for specified characteristics is $U_i = \frac{1}{n} \int u(\frac{dy}{dx})^{-2} \sum_{x\varepsilon C}^{n} P$…Eq.3

Resultant User Vector for all Characteristics $A_i = \int u(\frac{dy}{dx})^{-2} \sum_{x\varepsilon C}^{1} P$….Eq.4

User vectors on the user experience and behaviours have been estimated as subspace on multiple dimensions. Optimal Latent feature for the sentiment analysis is determined on utilizing scatter matrix[14] is provided by User Scatter Matrix on dynamic behavior $S_i = \int u(\frac{dy}{dx})^{-2} Ai$ ….Eq.5

Final Projection Matrix containing the user profile has been processed using transformation matrix through matrix normalization[15] on the selected user pool with similar characteristics [9]. The linear combination on task-adapted parameters of the user feature learning is carried out using pair wise similarities of the behaviour aspects. Latent Feature vector of the user has been obtained for user pool

3.3. Long Short Term Memory Recurrent Neural Network

LTSM is a important feature of RNN is to forecast or recommend the user adaption to the present application. In this work, user adaption rate to the e learning system is examined using the user profiles and user preference associated with it. RNN is employed to process sequence of the user intention. An LSTM architecture considered as integration of various states such as cell state, input layer, input gate, forget gate and output gate. It outcomes the directed acylic graph which produces the back propagation to the user association rate to the particular learning model. In RNN based LTSM, each layer of the LTSM network is stacked vertically. Table 1 illustrates the parameter of Multilayer RNN-LTSM

Table 1: Hyper Parameter Multilayer RNN-LTSM

Hyper Parameter	Values
Batch Size	168
Learning Rate	0.02
Number of Epoch	50
Number of data points	10000
Sequence length	500
Error function	Cross entropy

- **Abstraction layer**

In this layer, high level features of the user profile are extracted in the abstraction layer of the network. It utilizes the activation function to represent the feature extracted in the vector form. Vector is arranged as gradient decent on sequence of the user behaviour. Abstraction layer is capable of the identifying the next state of the user on analysing the present state with few functional parameters and epoch. It is processed by multiplying the user intention vector with the weight matrix. Further biased vector is termed as next state or adaptation rate of the user. It is given as

Adaptation Rate = (tanh)(W[S_i(u)] Eq.6

Where tanh is the activation function for computation of user adaption to the current aspect and adaption value is stored in the cell state

- **Hidden Layer**

In this layer, hidden vectors contain the user states. Further state represents the feature extracted in the abstract layer and incorporates the forget gate. The forget gate of LTSM Model is to compute the state information stored in the cell state with respect to other available data and those information has to be hidden and eliminated. The forget gate of LTSM employs the sigmoid function as its activation function for the identify the sentiment or opinion to be the collected hidden information. Figure 1 represents the architecture of the proposed model.

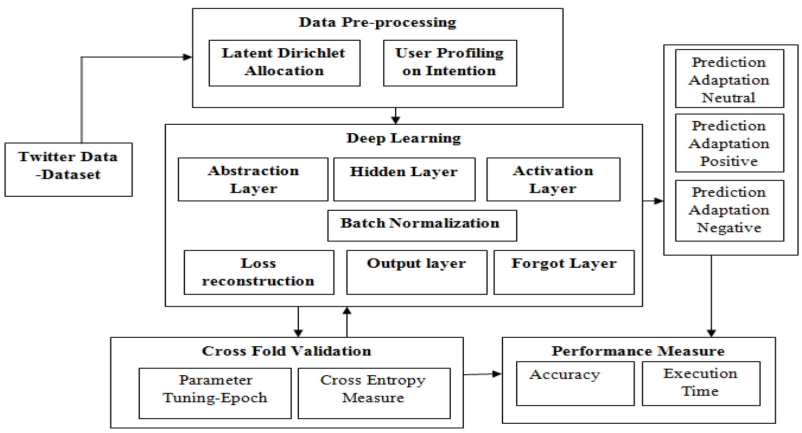

Figure 1: Architecture of proposed Forecasting Adaption Rate

Each user state in the LTSM layer is hidden
The output of the forget gate containing user intention is provided by
$gf(t) = \sigma(Wf\,[u(t); S(t-1)] + bias)$...Eq.7
Hidden User Vector is provided by $h_t = F_t\,(U_{t-1}\,U_t)$
$h_t = \tanh(w_h U_{t-1} + w_{xh} U_t)$...Eq.8

The historical information of the user data in the LSTM layer considered to be vital for the analysis of the sentiment to the intention stored in the user state u(t). The input gate of LTSM determines which parts of the input containing the adaption context are significant in storing for forecasting the future adaptation. Such decision masking is arrived by evaluating the information of current user intention u(t) and the previous user intention u (t − 1) together
$gi(t) = \sigma(Wi[c(t); z(t-1)] + di)$...Eq.9

Where gi(t) is the prediction output of the input gate, W_i and b_i are weight matrix and bias vector of the input gate representing the user intention on various context respectively, and $\sigma(\cdot)$ is the sigmoid activation function to enable processing of forget gate on the inputs from the input gate. The sigmoid activation function of the input gate provides the assurance that the output of the input gate $g_i(t)$ will been constrained in the range of sentiment(positive , Neutral, Negative) . However gi(t) near to the sentiment range allows the input information of user intention to the specific concept is to be stored in the user state of LTSM layer. .

- **Activation function**

Thus, the learning rate of the proposed model is to controls and adjusts the weights of the sequence of user intention using sigmoid function. The activation function of multilayer RNN-LTSM network is represented in many to one structure. The activation function is represented with bias to produce the output. It is provide as follows
$O_s = \tanh(U_s)$...Eq.10

It is employed for analyzing the user features as intention to identifying adaptation rate and expressing the features to represent their variations.

- **Output Layer**

The particular layer works as output gate of the LTSM model. The output gate of LTSM model computes which part of the information of the processed user state information should be utilized to produce its final output of the LSTM block. The activation function of the output gate go(t) is employed to filter the output containing the future prediction of the adaptation in terms of positive , neural and negative. Particular outcome is considered as resulting final output z(t).In current architecture, output layer predict the adaptation rate with respect to sentiment of user on basis of user intention , behaviours and perception..
$A(t) = (\tanh)(u(t)) * g_i(t))$...Eq.11

The hyperbolic tangent function of the LTSM is employed as the activation function of the output layer towards identifying the adaptation rate of user to e learning system by producing the output of the LSTM block in the range of sentiment (positive, Neutral , Negative), which also guarantees the network not to diverge for long sequence of input data of the inputs gate.

- **Loss Layer**

This layer is to ensure the prediction accuracy on resultant outcomes on fine tuning against refine parameter of various layers of recurrent neural network to ensure the minimum reconstruction error among hidden layer and sigmoid activation layer.
$l_t = \text{softmax}(U_w * U_t)$...Eq.12

Further cross entropy loss function has been employed to handle manage data seperability of prediction information. Model parameter of neural network is updated to predict the result of adaption. Soft max and cross entropy mechanism on the instances has been carried out to get effective prediction outcome to various domain

Algorithm 1: Prediction of adaption rate on basis of sentiment
Input: Twitter Dataset
Output: Adaption rate of user on basis of sentiment to application
Process
Data Pre- Process ()
User Profiling ()
Extraction Intention and Preference
Sentiment lexicon generation using LDA()
Apply Multilayer RNN_LTSM Learning ()

Abstract learning ()
Determine High Level Feature
Hidden Layer ()
Extract the latent user intention() and User preference() as states
Forget Layer ()
Compute eliminated intention()
Activation Layer ()
Use Sigmoid Function Tanh()
Cross Entrophy Layer ()
Minimize Reconstruction Error ()
Output Layer ()
Softmax ()
Output: Forecast User Adaption on Sentiment (Positive, Neutral, Negative)

Algorithm Description:
Initially algorithm processes with pre-processing technique to identify the user intention on basis of user preference and user behaviour profiling. Preprocessed data is used for prediction of the adaptation rate of the user to the application on basis of sentiments. In order to identify the sentiment, Feature extraction has been carried out on the dataset using Latent Dirichlet Analysis through generation of scatter matrix. The generated features is processed in the multilayer RNN-LTSM architecture to yield adaption rate with respect to the learning challenges and advantages provided in form of user intention on usage of abstract learning , hidden layers, activation layer and output layer[16].

4. Experimental Results
Experimental analysis of RNN-LTSM based learning architecture using parametric tuning on Twitter data for predicting adaption rate of the user with respect to sentiment has been carried out in this work. The performance analysis of the proposed model has been evaluated utilizing precision, recall and Fmeasure metrics. The proposed model is experimented and evaluated. Finally performance of the model is cross evaluated using 10 fold validation. The training parameter of the comprehensive learning has been illustrated in the table 2

Table 2: Training parameters

Parameter	Value
Learning rate	10^{-6}
Loss Function	cross entropy
Batch size	20
Epoch	50

4.1. Dataset Description
We have carried out extensive experiments on Twitter datasets composed of 30000+ tweets in order to measure the outcome of the students towards adaption of the application in future. In this model, dataset produces the data segmented into equal parts for training and testing. In this experiment, training of model consumes 60%, Validation consumes 20% and testing consumes 20%.

4.2. Evaluation
The proposed architecture for prediction of adaption rate of the user to application has been evaluated against the following performance measures against conventional machine learning approaches. In this work, proposed architecture is validated using 10 fold validation to determine the performance of prediction results with respect to time specific twitter data. The performance evaluation of the proposed multilayer RNN_LTSM model with on the process of activation function, Hidden layer, loss function and Output layer with softmax processing.

- **Precision**

Precision is a termed as measure of Positive predictive value of the sentiment class. It is further depicted as the fraction of similar instances among the each class groups generated using the model. Figure 2 illustrate the performance evaluation of the proposed architecture in terms of precision on dataset. Performance measures are suitable for determining the feasibility of proposed architecture on adaption rate prediction. Effectiveness is achieved due to hyper parameter tuning

$$\text{Precision} = \frac{True\ positive}{True\ positive + False\ Positive} \quad .....\text{Eq.13}$$

Precision is computed using True positive and true negative. True positive is a number of similar points in the pattern generated and false negative is number of real dissimilar points in the pattern[15]. Mostly a good performance is also characterized by Euclidean distance[16] for the data points. It can be calculated using recall measure.

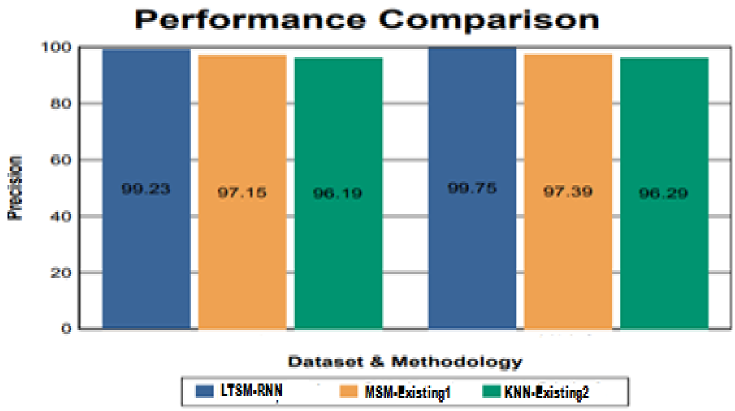

Figure 2: Performance analysis of the methodology on aspect of Precision.

• **Recall**

Recall is the part of similar instance of the dataset that have been extracted over the total amount of relevant instance of the dataset[17]. The recall is the part of the similar documents that are successfully classified into the exact classes.

$$\text{Recall} = \frac{True\ positive}{True\ positive + False\ negative} \text{ ...Eq.14}$$

True positive is a number of similar data points in the pattern and false negative is number of similar data points in the pattern extracted. Figure 3 represents the performance evaluation of the proposed architecture on recall measure along state of art approaches.

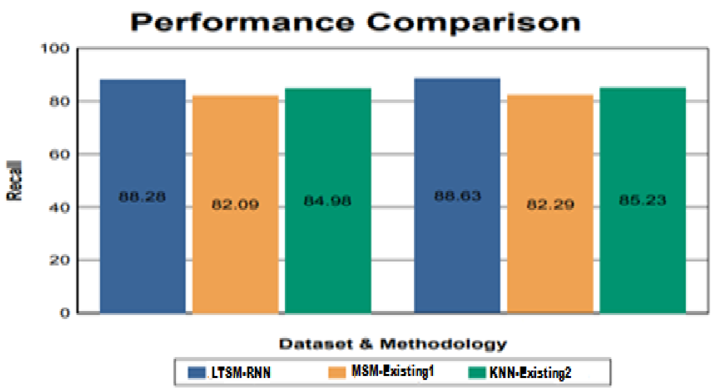

Figure 3: Performance analysis of the methodology on aspect of Recall

Prediction quality depends on activation function in every layer. Forget layer calculates the feature map to generate subspace. F measure is a good measure for determining the quality of the clustering.

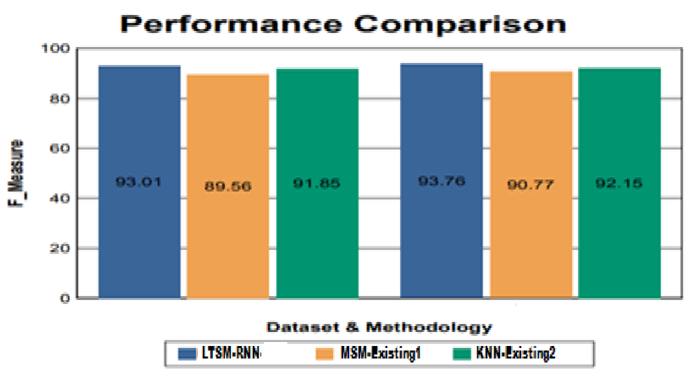

Figure 4: Performance analysis of the methodology on aspect of F- Measure

• **F measure**

It is the number of correct class predictions to the incoming data among total number of predictions to whole category of data[18].

Accuracy is given by

$$\frac{True\ positive + True\ Negative}{True\ positive + True\ Negative + false\ positive + False\ negative} \text{...Eq.15}$$

The different user intention generated may have different adaption rate. Figure 4 represents the performance of the proposed model in terms of f measure against state of art approaches for adaption rate formation with respect to sentiments[19]. Table 2 presents the performance value of the technique for sentiment analysis.

Table 2: Performance Analysis of Proposed architecture against state of art approaches to EHR dataset – Category 1

Technique	Precision	Recall	F measure
LTSM-RNN Learning- Proposed	0.9923	0.8828	0.9301
Multiscale Matching based Classification Learning –Existing 1	0.9715	0.8209	0.8956
KNN Based Learning - Existing 2	0.9619	0.84.98	0.9185

The prediction of the adaption with sentiment on the dataset is made effective on using LTSM optimization to results obtained. The model is capable of detecting the underlying structure of the data distribution in general. Hyper-parameter tuning [20] is a very important component of the proposed models. In addition, the cross-validation has been used to dataset alone to compute the good value of the hyper-parameters on sentiment analysis.

Conclusion

We designed and implemented RNN-LTSM technique for Sentiment analysis of the e learning data from twitter dataset. Proposed model utilizes the LTSM-RNN based parameterized and loss function for high representative prediction of the adaption rate of the user to the application on basis of sentiment attached to user intention.. Deep learning architecture utilizes the features in Hidden pooling layer to represent sparse feature for Sigmoid activation function and avoid the sparse feature in forget layer. Further softmax layer and loss layer has been embedded to produce the highly discriminative prediction patterns. Finally generated prediction outcomes have been optimized using LTSM technique to generate optimal sentiment to their adaption rate. Prediction performance computed using f measure proves that it is effective on prediction result on parameter optimization.

References

1. Min E, Guo X, Qiang "A survey of clustering with deep learning: from the perspective of network architecture" in IEEE Access, Vol. 6, issue.39, pp: 501–14, 2018.
2. Chowdary NS, Prasanna DS, Sudhakar P. "Evaluating and analyzing clusters in data mining using different algorithms". International Journal of Computer Science and Mobile Computing, Vol.3, PP: 86–99, 2014.
3. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1798–1828, 2013
4. Dizaji KG, Herandi A, Cheng," Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization. In: 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017, 5747–56.
5. Xie J, Girshick R, Farhadi A "Unsupervised deep embedding for clustering analysis. In: International Conference on Machine Learning. New York City, NY, USA: ICMLR, 2016, 478–87.
6. S.Praveen & R.Priya " A Deep Conceptual Incremental learning Based High Dimensional Data Clustering model- A Deep Learning Approach" in Turkish Journal of Computer and Mathematics, 2021
7. H. Liu, M. Shao, S. Li, and Y. Fu, "Infinite ensemble for image clustering," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, , pp. 1745–1754, 2016
8. J. L. Elman, "Distributed representations, simple recurrent networks, and grammatical structure," Machine learning, vol. 7, no. 2-3, pp. 195–225,1991.122, 2015.
9. K. G. Dizaji, A. Herandi, C. Deng, W. Cai, and H. Huang, "Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization," in 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, , pp. 5747–5756, 2017.
10. P. Xiao, G. K. Venayagamoorthy, and K. A. Corzine, "Combined training of recurrent neural networks with particle swarm optimization and backpropagation algorithms for impedance identification," in Proceedings of the IEEE Swarm Intelligence Symposium (SIS '07), pp. 9–15, April 2007.
11. P. Huang,Y. Huang,W.Wang, and L.Wang, ``Deep embedding network for clustering," in Proc. 22nd International. Conference. Pattern Recognition. (ICPR), Aug. 2014, pp. 1532-1537.
12. S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," Med. Image Anal., vol. 16, no. 7, pp. 1385–1396, Oct. 2012.

13. W. Harchaoui, P. A. Mattei, and C. Bouveyron, ``Deep adversarial Gaussian mixture auto-encoder for clustering," in Proc. ICLR, 2017, pp. 1-5.
14. N. Dilokthanakul et al. (2016). ``Deep unsupervised clustering with Gaussian mixture variational autoencoders." [Online]. Available: https:// arxiv.org/abs/1611.02648
15. G. Chen. (2015). ``Deep learning with nonparametric clustering." [Online]. Available: https://arxiv.org/abs/1501.03084
16. S.Praveen & R.Priya "A unified deep learning framework for text data mining using deep adaptive fuzzy clustering" in Solid state technology, Vol.63, issue.6, 2020.
17. J. Chung, Ç. Gülçehre, K. Cho and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," 2014. [Online]. Available: https://arxiv.org/abs/1412.3555. [Accessed 31 1 2019].
18. T. Pham, T. Tran, D. Phung and S. Venkatesh, "Predicting healthcare trajectories from medical records: A deep learning approach," J. Biomed. Inform., vol. 69, pp. 218-229, 2017.
19. K. Zhou, H. Zha and L. Song, "Learning triggering kernels for multidimensional hawkes processes," in Proc. 30th Int. Conf. Machine learning, Atlanta, GA, 2013, pp. III-1301-III-1309.
20. T. Ma, C. Xiao and F. Wang, "Health-ATM: A deep architecture for multifaceted patient health record representation and risk prediction," in Proc. 2018 SIAM Int. Conf. Data Mining, San Diego, CA, pp. 261-269