How to Cite:
Kumar, K. A., & Vasundra, S. (2022). A novel concept specific deep learning for disease treatment prediction on patient trajectory data model. *International Journal of Health Sciences*, *6*(S1), 13135–13148. https://doi.org/10.53730/ijhs.v6nS1.8290

A novel concept specific deep learning for disease treatment prediction on patient trajectory data model

K. Arun Kumar*
Research Scholar, Dept. of CSE, JNTUA, Ananthapuramu - 515002, India.
*Corresponding author

S. Vasundra
Professor, Dept of CSE, JNTUACEA, Ananthapuramu - 515002, India.

***Abstract*---**Prediction of diseases by researchers in healthcare industries and the different machine learning algorithms involved play an important role in disease prediction. Similarly, the prediction of patient trajectory data and its analysis also plays a vital role. A novel deep learning approach named Deep Concept Specific learning has been proposed for analyzing unstructured patient trajectory data. The proposed model is structured as a generic model suitable for any patient's medical disease. To evaluate the proposed model performance, the disease related to diabetes has been employed. It implies for autoencoder model, which has been optimized with hyper parameter tuning. It effectively transforms high dimensional features into low dimensional feature space to extract concept-specific features. It is implemented on the Further, all the parameters employed in the sparse features are manipulated. They are considered a loss and encoder function to compute the optimal features for effectively predicting the results. The encoder function maps the data points containing optimal features into latent representations based on disease characteristics. It is done from the patients' perspectives by the ReLu activation function based on generated feature characteristics. Extensive experiments have been performed on real datasets such as the UK CF trust and Mimic-III datasets to compare the proposed model with popular state-of-the-art approaches. The experimental results show that the Deep Concept specific clustering can achieve both effectiveness and good scalability with an accuracy of 98.25% against the various state-of-the-art techniques.

***Keywords*---**Patient Trajectory Data, Deep Learning, Data Distribution, Unstructured data, Reconstruction Error

Manuscript submitted: 18 March 2022, Manuscript revised: 9 April 2022, Accepted for publication: 27 May 2022

Introduction

Prediction is a primary aspect of unsupervised learning representation employed in exploratory data mining applications, especially in the medical domain. The objective of patient treatment Prediction is to categorize distributed treatment trajectory data points into one or more patient classes based on similarity measures on the data points. However, a large extent of disease data classification methods is employed in various categories of a dataset that has been presented [1]. Traditional data classification models usually have exhibited limited performance on patient trajectory data as it is unstructured [2] and it is incapable of estimating the prediction Results through computation of similarity measures. Furthermore, those data classification methods generally degrade in particular aspects like the curse of dimensionality, data sparsity, and computational complexity on employing it in large-scale datasets [3][4]. To tackle those implications, dimensionality reduction and feature transformation methods [5] like Principal Component Analysis (PCA) [6] and Linear and Nonlinear Transformation are named as Kernel methods [7], and Spectral Methods [8] have been utilized. This was implemented on Artificial Neural Network and decision tree algorithm for patient trajectory data management in more significant aspect by mapping the input data into a new text feature space for data classification. Nevertheless, a large composite latent structure of text data is highly challenging against its effect on state-of-the-art classification methods. To improve text classification performance, learning representatives based on deep learning approaches can be developed to transform the trajectory data into more class-specific data distributions.

Deep learning architecture for prediction has been employed using neural networks for better representation of patient trajectory data, which is represented in an unstructured format. This article proposes deep concept-specific clustering as the deep learning approach for analyzing unstructured high-dimensional patient trajectory data containing patient treatment information, disease diagnosis information, and admission and discharges information. The approach uses an auto encoder category to obtain the feasible feature space containing concept-specific features. The concept-specific feature construction has been carried out with the wordnet tool and pos (prognosis outcomes) tagging process in the autoencoder layers. Further optimization of the deep layer for concept-specific feature employed with hyperparameter tuning of epoch and bias weight parameters. The optimized Auto Encoder model provides a nonlinear mapping function to construct the sparse features with loss criteria to include the latent representation of the unstructured data into clusters. Eventually, learning of composite data representatives in the form of clusters through mappings will allow data transforming into more generic data clustering representations as the streaming data has been mapped to a low dimensional feature space. Treatment Prediction to the appropriate cluster has been carried out using distance computations.

The paper is organized as follows, related work of the prediction models are described in section 2, the architecture of the proposed deep concept specific learning approach is described in section 3 and experimental results and proposed model effectiveness is demonstrated in section 4 using a dataset on basis of experimental comparison with the state of arts existing approaches on various performance measures has been explained. Finally, the article has been concluded in section 5.

Related Work

In this part, the data classification model using machine learning approaches has been evaluated in detail based on machine learning architectures for feature representations and similarity measures of the clustered data points and problem statements.

Problem statement

Patient Treatment Trajectory data is used to predict the outcome to determine the Survival rate and the response of the patient to the treatment. Trajectory data has been limited to only specified disease prediction on employment with high accuracy and less computation cost. The medical community is trying to reproduce that generic model based on various health-related factors, social factors, and Lifestyle attributes of the patient to improve the efficacy of medical treatment. Furthermore, the novel solution required to detect any disease and its treatment procedures for detectable changes in the diseases based on various factors of patients.

Review of Literatures

Various works of literature related to patient trajectory data mining model for treatment planning against various aspects have been highlighted.

- Nenad Tomasev [9] et al. proposed an experimental model that uses Hubness's role for large-scale data clustering to distinguish distances between data points and observe a lower-dimensional feature subspace. It contains points (hubs) that periodically occur in k-nearest- a neighbor outcome with similar data points. It can be accurately utilized for clustering of those data points. It can be used effectively as a clustering prototype for large-scale mining data for centroids-based cluster configurations.
- Sangdi[10] Lin et al. proposed an experimental model that uses the CRAFTER, which can simultaneously manage the data based on categorical distribution attributes and numeric related attributes on various aspects. It is capable of scaling any dimensionality the size of datasets. The concept of the data class probability is incorporated to determine the cluster representatives as a group containing data points on the clustering process. Cluster is more sensitive to irrelevant attributes, data noise, and class outliers, and it is suitable for any size of the dataset.
- Punit Rathore[11] et al. proposed a model that uses the Rapid form of the Hybrid Clustering algorithm, which is employed for clustering large velocity and volumes of a high dimensional dataset. It is to differentiate the data cluster distributions in the dataset collected using important three sampling strategies of the clustering: random sampling, MMRS based data sampling in the p dimensional up space, and MMRS based data sampling in the q dimensional down space.
- C.C. Aggarwal[12] et al. proposed a model that uses the generalized projected

clustering technique for large volume datasets, which may also be viewed as a way of planning reconfiguring the clustering for large-scale dimensional applications by identifying hidden subspaces with clusters that are determined by inter-attribute correlations.

- Kaban[13]. et al. proposed a model which employs Non-Parametric Clustering of High Dimensional Data Detection on Meaningless Distances computation using phenomena of large scale dimensional data processing, subspace analysis, data retrieval, and data point indexing, which all rely on distance-based clustering category of distance or dissimilarity of the cluster data points. It uses the finite-dimensional characterization of the
- Data instance distance specific concentration phenomenon that reorganizes last obtained output results in the restrictions of infinite data dimensions of the input.
- N. M. Nawi[16]. et al. proposed a back-propagation neural network for predicting patients with heart diseases. Further obtained cluster is classified and predicted based on gradient descent optimization. It enhances the computational efficiency and convergence rate of the data prediction.

Analysis of various data mining approaches towards predicting disease treatment of a patient

In this section, various machine learning and deep learning approaches have been analyzed to a different category of the diseases against multiple factors to build treatment prediction on patient trajectory data listed in table 1.

Table 1: Approaches to predict disease treatment of the patients

SI.NO	Title	Objective	Advantages	Disadvantages
1	Hubness Phenomena	It observes the lower-dimensional feature space to predict the treatment to any disease	It clusters the dataon various time intervals effectively	Scalability is low
2	CHAPTER technique	It correlates both categorical attributes andnumerical attributes on various aspects	It eliminates the noise and outliers of the clusters	Computation timeis high
3	Rapid Hybrid Clusteringalgorithm	It clusters large volume and high data changing the instances	It achieves high accuracy	It increases the outfitting issues
4	Projected Clustering	It identifies the hidden subspaces in the cluster	It eliminates the overfitting issues	It leads to NP-hard Problem
5	Back Propagation Neural Network	It Computes the gradient descent in terms of weight	Accuracy is high	Processing of the model is very slow

Proposed Model

This section provides an informal definition of the distributed data classification approach and later presents the deep concept-specific learning framework for mining evolving patient trajectory data streams.

Data Pre-processing

A large variety of datasets of a form of patient trajectory data are unrealistic for processing. Data Pre-processing has been applied in missing value prediction and dimensionality reduction.

Missing Value Imputation

Missing Value Imputation has been used in the concept of factor analysis. Factor Analysis determines maximum common variance on the particular data field. It follows the Kaiser criterion, which uses the eigenvalue. It uses the score for the variance of the specific data field to fill the missed value of the data field. It can also compute using the maximum likelihood method based on the correlation of the data field [10].

Dimensionality Reduction

The dimensionality reduction technique uses Principle Component Analysis (PCA) to reduce the input data consisting of high dimensional data to lower dimensional data. It is further used to eliminate overfitting. PCA is a linear transformation technique. It reduces the feature based on correlation. It aims to project the subspace with fewer dimensions in high-dimensional data. It has been processed using a dimensional transformation matrix [11].

Feature Selection

Feature selection of deep learning uses multiple layers to extract the features. The features are extracted from the hidden and deepest layers [12]. It can obtain accurate feature representations from the streaming data to increase the separation of obtained data instances of the cluster using similarity computation. It learns hidden features in the pre-processed data. It maps it to the encoder layer of the model to encode and decode the data in the layers without the inclusion of the probability distribution function on the input selected samples, as represented in equation 1 as follows.

Given a set of data points of p variables $\{x_1, x_2, \cdot, x_n\}$ ....(Eq.1)
The optimum data mapping on layer based objective criterion is given in equation 2 for y=f(x) as Vector is given by

$$\mathbf{x} = \sum_{i=1}^{N} x_i v_i = x_1 v_1 + x_2 v_2 + \ldots + x_N v_N \quad \text{................................ (Eq.2)}$$

The low-dimensional vector is represented as

. $$\hat{\mathbf{x}} = \sum_{i=1}^{K} y_i u_i = y_1 u_1 + y_2 u_2 + \ldots + y_K u_K \qquad \ldots\ldots\ldots\ldots\ldots\ldots \text{(Eq.3)}$$

The above representation of the vector containing data set generates the feature vector o n a particular objective function to produce the maximum discriminatory information as represented in equation 3

Deep Concept Specific Clustering

Deep Concept Specific Clustering has been employed for clustering the feature extracted in a hidden layer of proposed optimized hyperparameter tuned autoencoder architecture. In this part, the Auto encoder-based deep learning model has been used. Initially, the encoder function has been employed to map the extracted features into concept-specific features representation based on using Term Frequency Computation, Parts of Speech tagging. Concept learning of the components using WordNet tool and these representations have been processed in decoder function to reconstruct features based on various disease characteristics have been partitioned into clusters [13]. The encoder and decoder function has been constructed as a fully connected neural network. The encoder and decoder function is represented as

Φ: feature spaces Fc → Ed

Where E_d is encoder outcome which is transformed as F ε R^pΨ: R^p Encoder Form → L_c(h)

Where L_C is the class of the features containing the latent representation of the features

The optimization objective function composed of a hyperparameter used in the fully connected neural network composing disease category significant features is given in equation 4 as

C = λLc + (1 − λ)Lc ……. (Eq.4)

Where λ is considered as a hyper parameter and L_C is considered as a class limit.

This function appreciates the feature points on a representative map to form a cluster or achieve more discriminative features in the particular cluster limit. The class function is optimized based on the weight function specific to disease and treatment and its bias.

X= ϭ(wH+b) ……. (Eq.5)

Where b is the bias function to map the disease to the treatment prediction on basis of patient

The function minimizes the reconstruction error on the average training set of the feature space consisting of the feature vector in the compressed form in t h e encoding and decoding process. It identifies the treatment for any specific disease using identity function. Identify function uses the treatment vector for diseases as augment for diseases. Figure 1 represents the architecture of the proposed model.

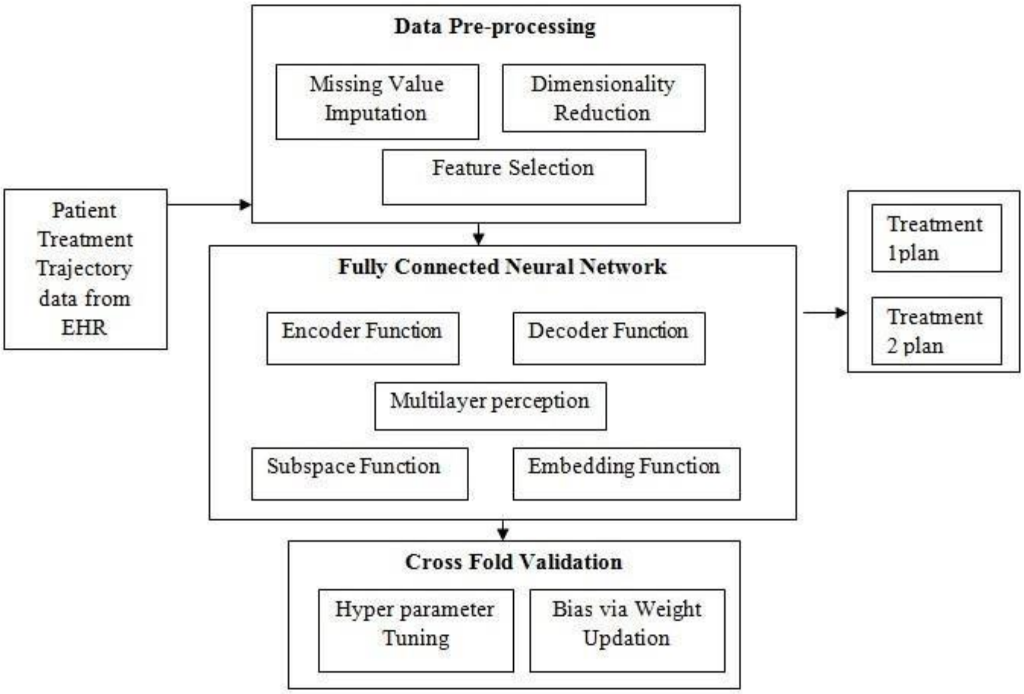


Figure 1: Architecture diagram of the proposed deep concept specific learning

Figure 1 represents the architecture of the proposed model with functionalities of the auto encoder model to process the patient diagnosis data with novel data managing mechanisms such as concept-specific learning for the treatment trajectory information. The proposed model extracts the latent feature on deep learning layers on generation effective feature subspace patients' diagnosis information for mapping. It is effective in mapping multiple diseases. The obtained features are classified based on the multiple factors of the patients. Obtained class of the instances represents high similarity on distance calculations on hyperparameter tuning in cross-fold validation of the dataset to evaluate the proposed model.

Hyperparameter Tuning

A hyperparameter is essential for balancing the classes' data instances composed of many overfitting features. In this, Variants of autoencoder on different perspectives have been used for the optimization of hyperparameter on the activation function, which is represented in Table 2

Table 2: Hyperparameter value for the embedded layer

Hyper Parameter	Values
Batch Size	148
Learning Rate	0.05
Size of word vector	250
Loss function	Cross entropy

Embedding Function

The embedding function is imposed in the hidden layer of the autoencoder model to minimize the reconstruction error. Dimensions are embedded in each data point with D into a d-dimensional space. Simultaneously, it is required that the data points in the hidden space can be recovered to their original form, as given in equation 6

$$S_w = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (x_{ij} - \boldsymbol{\mu}_i)(x_{ij} - \boldsymbol{\mu}_i)^T \quad \text{..............................(Eq.6)}$$

In other words, the auto encoder is used here as a complementary regularize to the embedding process. Hopefully, we can derive the embedding with more semantic representations by this joint minimization.

Subspace Function

Especially transfer learning uses the three constraints; initially, it captures deep autoencoder functions to learn the reduced feature representation of the input from the streamed data to select the feature instance as subspaces. It is to capture the extra features of the data based on unstructured patient data. The following process preserves the data structure to conserve the local structure of data properties of the original data by employing a locality saving constraint that has been applied on the subspaces composed of extra features [15]. Additional features of the dataset contain the patients' admission information and discharge information. Finally, it also includes a subspace instance sparsity constraint to determine the data affinity of feature representations.

Multilayer Perception

In this process, weights for input have been applied to features x1, x2, x3... xn. However, updating will concentrate on nets with units arranged in layers with a learning rate γ. It is multiplied by weight[16]. The multilayer perception is given in equation 7 as toa= x1w1+x2w2+x3w3... +xnwn

Each Unit in perception is represented as $y_i = f(\sum_{j=1}^{m} w_{ij} x_j + b_i)$ ....(Eq.7)

Perceptron eliminates the problem related to the sigmoid. Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result of the treatment prediction.

Bias Weight Updation

The autoencoder uses the word embedding as an embedding function of a neural network to find the concept of the feature to map into the cluster. Further, it learns the nonlinear dependency of the data points and uses transfer learning on the encoder. Finally, it eliminates the data sparsity issues effectiveness on weight updating through bias function [14]. Bias weight updating is given in equation 8 as

$$\varepsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2n}} \quad \text{...(Eq.8)}$$

The function between the input and hidden layers of the autoencoder has been mapped to treatment prediction. Without accessing the original training data set, it can directly map an instance into the low-dimensional space. Thus, it supports incremental embedding more efficiently.

Algorithm 1: Deep Concept Specific Learning Input: Patient Treatment Trajectory Data D

D=[d1,d2,d3....dn}

Output: Classes C={c1,c2,c3..Cn}Process

1. Data Preprocessing ()
2. V(d) -> Vector form
3. Compute Missing Value ()
4. Centroids K for v[d]
5. Assign K for missing Field F
6. Feature Reduction ()
7. Dimensionality Reduction _PCA()
8. Concert Vector V(d) to Matrix m(d)
9. Compute eigenvalue for M(d)
10. Compute Eigen Vector for M(d)
11. Eigen vector E(v) is FS
12. Return FS of the patient disease
13. Model Deep Concept Specific learning ()
14. Generate Subspace SB for FS
15. Calculate Latent Features LS on SB
16. Latent Feature of Treatment Trajectory
17. Transfer learning ()
18. Encode SB ()
19. Map latent LS with FS20. Decode ()
21. Bias a Map LS
22. Compute distance of the FS and LS
23. Group FS w.r.t distance f_i and $f_{i-1} < T$
24. Classes C= {c1,c2,c3..Cn} for treatment for the disease

The algorithm of deep specific learning has been applied to generate the class which maximizes the accuracy and minimizes the reconstruction error for patient data related to diabetics. Classification of the cluster is done in the softmax layer which uses distance measures to categorize the patients' response to the medication. It is done with the support of admission and discharge information of the data features.

Hyper Parameter Tuning

Hyper parameter tuning is carried out on layers of the model containing the sparse features. In this work, Bias weight and epoch parameter has been used for tuning. These parameter increases the weight of the feature to eliminate the overfitting feature instance to the class on processing on features obtained hidden layer. However, measuring each of the data points in hyperparameter tuning[17] would be complicated in this comparative analysis of the prediction model. Hence, details of bias and weight updates as a training model for the AE methods have been detailed through epoch, and Cross-validation on test data includes every possible step represented in proposed approaches. Finally, disease treatment to the patient is predicted based on similarity of admission features instances with other discharge patients with the same admission feature values in the test process.

Experimental Results

Experimental analysis of proposed deep learning architecture for cluster generation has been carried out on the UK CF trust and Mimic-III dataset, which is high dimensional, containing 124789 data instances of patient trajectory information. The performance of the proposed technique has been evaluated using Precision, recall, and F-measure. Further, the model's performance is cross-evaluated using ten-fold validation. Figure 2 represents the performance evaluation of the architecture in terms of Precision on the electronic health care dataset.

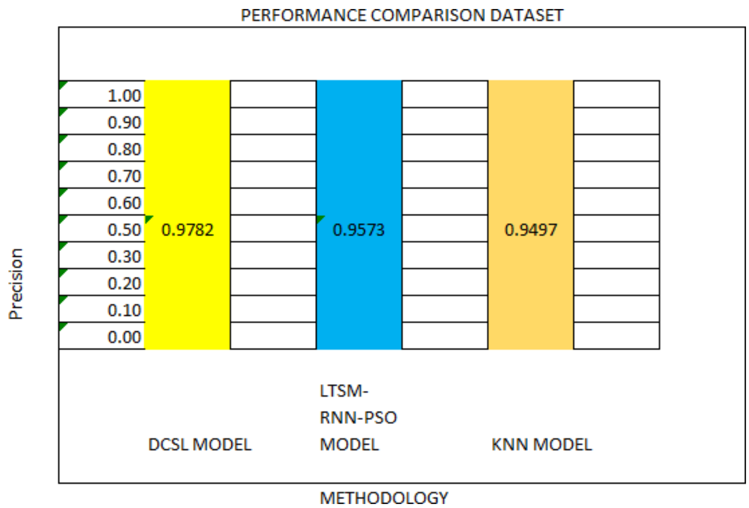


Figure 2: Performance analysis of the methodology on an aspect of Precision

The computation of precision is to determine the prediction efficiency of datamanagement. It ismeasured as no of relevant records retrieved on the no of instances.

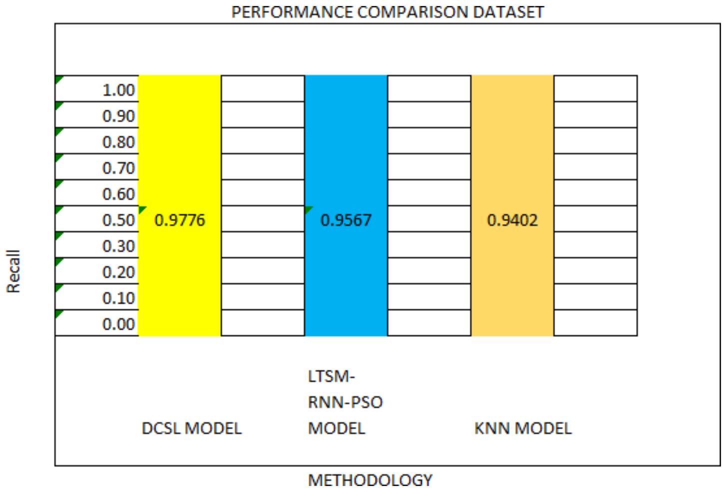
$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{FalsePositive}} \quad \text{............(Eq.9)}$$

Precision is computed using True positive and true negative. True positive is similar points in the pattern generated, and false negative is real dissimilar points [18]. Performance measures are suitable for determining the feasibility of the proposed architecture on class generation for diabetic disease. Effectiveness is achieved due to hyper-parameter tuning. A recall measures the number of correct relevant instances on the dataset. Recall estimate is used to identify the prediction accuracy. The recall is the part of the similar documents that are successfully classified into the same classes.

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \quad \text{........................(Eq.10)}$$

True positive is similar data points in the pattern, and false negative is similar data points in the pattern extracted. An effective data clustering performance has been characterized by measures of intra data cluster similarity and inter-data cluster similarity for the data points. It can be calculated using recall measures.

Cluster quality depends on the activation function in every layer. The encoder calculates the feature map to generate subspace. F measure is an effective performance measure to compute the performance of the data clustering. (Figure 3) represents the performance evaluation of the proposed architecture on recall



measure and state-of-the-art approaches.

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

It is the number of correct class predictions to the incoming data among a total numberofpredictions to the whole category of data.

$$\text{F measure} = \frac{\text{True positive} + \text{TrueNegative}}{\text{True positive} + \text{True Negative} + \text{false positive} + \text{False negative}} \quad \text{........(Eq.11)}$$

Figure 3 represents the performance of the proposed architecture in terms of f measureagainst state of art approaches for patient treatment prediction.

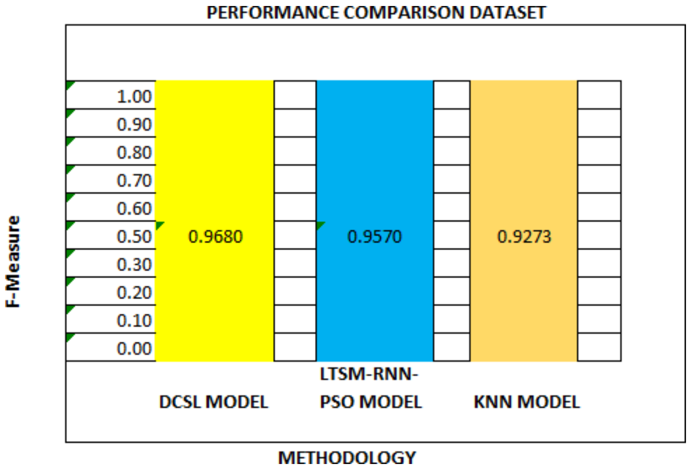


Figure 4: Performance analysis of the methodology on the aspect of F- Measure

The proposed encoder function is an approximation objective function to map the data input into a cluster distribution of the particular cluster [19]. Then, the generative probabilistic decoder function of the proposed model of the patient trajectory model tries to produce accurate prognosis outcomes through the data conditional probability function. Further hyper parameter tuning based on epoch and bias weight produces effective classification and prediction results. Table 3 presents the performance value of the technique for cluster analysis on eliminating the overfitting issues of the data analysis through hyper-parameter tuning.

Table 3: Performance Analysis of Deep learning architecture against state of art approaches

Technique	Precision	Recall	F measure	Accuracy
Deep Concept Specific Learningmodel – Proposed	0.9782	0.9776	0.9680	0.9921
LTSM-RNN-PSO Existing 1	0.9497	0.9402	0.9273	0.9448
K-NN based learning – Existing 2	0.9619	0.8498	0.9185	0.9234

On the other hand, the prediction methods are not only for detecting disease prognosis on the patient trajectory data but also for predicting the underlying cluster structure of the feature data distribution in general for concept-specific learning to generate a generic model. It is much more powerful than existing approaches to handling the patient trajectory data. Injecting auto encoder regularization [20] may help guide the embedding towards better data representation to predict the treatment of any diseases.

Conclusion

Deep Concept Specific classification is a deep learning technique designed and implemented for analyzing unstructured Patient Trajectory data. The proposed model uses the autoencoder-based fully connected neural network via generating the cluster with minimized reconstruction error. The model uses word embedding for concept-specific feature generation towards mapping the sparse representation into classes. Further hidden layer and softmax layer produces effective results for generic model towards disease treatment prediction. The generic model processed the various factors of the patient. Prediction performance computed using F measure proves that it is effective on cluster instance similarity by better initializing the parameters. Finally proposed model demonstrates that it is effective and highly scalable on patient Trajectory data.

References

Min, E., Guo, X., Qiang, E.,2018.A survey of clustering with deep learning: from the perspectiveof network architecture. IEEE Access, Vol.6, Issue.39 pp:501–514.

Chowdary, N.S., Prasanna D.S., Sudhakar, P.,2014. Evaluating and analyzing clusters in data mining using different algorithms. International Journal Computer Science in Mobile Computing, Vol.3, PP:86–99.

Davidson, I., Ravi, S.S.,2015.Agglomerative hierarchical clustering with constraints: theoretical and empirical results. European Conference on Principles of Data Mining and Knowledge Discovery.,Heidelberg, Germany, Springer. PP: 59–70.

Dizaji,K.G.,Herandi,A.,Cheng,2017.Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization.IEEE International Conference on Computer Vision (ICCV), Venice, Italy:PP: 5747–56.

Xie,J.,Girshick,R.,Farhadi,A.,2016.Unsupervised deep embedding for clustering analysis. In: International Conference on Machine Learning. New York City, NY, USA.PP:478–87.

Buza, K., Nanopoulos, A., Schmidt-Thieme, L.,2011.INSIGHT: Efficient and Effective Instance Selection for Time-Series Classification. Pacific-Asia Conference proceedings in knowledgeDiscovery and Data Mining (PAKDD), Part II, pp. 149-160

Parsons,L.,Haque,E.,Liu,H.,2014.Subspace clustering for high dimensional data: a review.ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 90–105.

Tomasev,N.,Radovanovic,M.,Mladenic,D.,Ivanovic,L.,2014. The role of hubness in clustering high-dimensional data. IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 3, pp. 739–751.

Kremer,H.,Kranen,P.,Jansen,T.,Seidl,T.,BifetA.,Holmes,G.,Pfahringer,B.,2011. An effective evaluation measure for clustering on evolving data streams, in Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 868–876.

Huang,P.,Huang,Y.,Wang,W.,Wang,L.,2014.Deep embedding network for clustering. In 22nd International Conference proceedings in Pattern Recount. (ICPR). pp. 1532-1537.

Ji,P.,Zhang,T.,Li,H.,Salzmann,M.,Reid,I.,2017.Deep subspace clustering networks. In Conference Proceeding of Advance Neural Network Information.Process and Systems. pp. 23-32.

Harchaoui, W., Mattei, P, A., Bouveyron, C., 2017.Deep adversarial Gaussian mixture auto- encoder for clustering. In Proceeding of ICLR. pp. 1-5.

Dilokthanakul, N.,2016.Deep unsupervised clustering with Gaussian mixture variational autoencoders.'' [Online]. Available: HTTPS:// arxiv.org/abs/1611.02648

Chen, G., (2015). Deep learning with nonparametric clustering. [Online]. Available: https://arxiv.org/abs/1501.03084

Jiang,Z., Zheng,Y.,Tan,H.,Tang,B.,Zhou,H.,(2016).Variational deep embedding: An unsupervised and generative approach to clustering.' [Online]. Available: https://arxiv.org/abs/1611.05148

Nawi,N.M.,Ghazali,R.,Salleh,M,N,M.,2010.The development of improved back- propagation neural networks algorithm for predicting patients with heart disease. In Information Computing and Applications: Proceedings of the 1st International Conference, ICICA, Tangshan, China, vol. 6377 of Lecture Notes in Computer Science, Springer, Berlin, Germany pp. 317–324.

Yan,Y.,Liu,Y.,Shyu,M.L.,Chen,M.,2014. Utilizing concept correlations for effective imbalanced data classification. In Proceedings of IEEE 15th International conference of Information Reuse Integration (IRI), pp. 561–568.

Chawla,N,V.,Bowyer,K,W.,Hall,L,O.,Kegelmeyer,W,P.,SMOTE: Synthetic minority over- sampling technique, Journal of Artificial Intelligence in. Resource. vol. 16, no. 1, pp. 321- 357.

Wang, H., Abraham, Z.,2015. Concept drift detection for streaming data" In Proceeding International Joint Conference of Neural Network (IJCNN), pp. 1-9.

Brzezinski,B.,Stefanowski,J.,2014.Combining block-based and online methods in learning ensembles from concept drifting data streams,' Information Science., vol. 265, pp. 50-67