SSRG International Journal of Computer Science and Engineering
ISSN: 2348 – 8387 / https://doi.org/10.14445/23488387/IJCSE-V10I5P101

Volume 10 Issue 5, 1-9, May 2023

Original Article

Fractional-Iterative BiLSTM Classifier : A Novel Approach to Predicting Student Attrition in Digital Academia

Gaurav Anand[1], Sharda Kumari[2], Ravi Pulle[3]

[1]*Senior Principal Software Engineer, WI, USA*
[2]*Systems Architect, CA, USA*
[3]*Principal Member of Technical Staff, CA, USA*

Received: 26 March 2022 Revised: 03 May 2022 Accepted: 15 May 2023 Published: 26 May 2023

Abstract *- Virtual learning circumstances have been observed as consistent growth over the years. The widespread use of online learning leads to an emerging amount of enrollments, also from pupils who have quit the education scheme previously. However, it also earned an increased amount of withdrawal rate when compared to conventional classrooms. Quick identification of pupils is a difficult issue that can be alleviated with the help of previous models for data evaluation and machine learning. In this research, a fractional-Iterative BiLSTM is used for predicting the student's dropout from online courses with a high accuracy rate. The feature extraction is provided by utilizing the encoder layer that efficiently extracts the features based on Statistical features. The Fractional-Iterative BiLSTM classifier is employed in the decoder layer, which is effectively performed in the classification function to predict the student dropout. The accomplishment of the research is evaluated by calculating the enhancement, and the developed model achieved the increment of 96.71% accuracy, 95.31% sensitivity, and 97.01% specificity, which shows the method's efficiency, and the MSE is reduced by 0.11%.*

Keywords *- Student dropout, Online learning, Fractional-iterative BiLSTM, Encoder layer, Decoder layer.*

1. Introduction

An appropriate development has recently been observed in virtual learning and online degrees. According to the statement provided by reliable Origins, virtual learning valuation is at $ 107 billion in the year 2015, which is predicted to obtain an impressive $ 325 billion by the year 2025 [1]. The reasons for this success are useful learning techniques, which arrogate the strategy of passing the limits of time and space to allow an essential amount of pupils to get admitted every time and everywhere in these classes [1]. Amidst various matters, the expense of virtual classes is highly attainable to every individual as well as organizations. Furthermore, virtual classes are very efficient when compared to conventional classes in providing knowledge to mature pupils coming to institutions after taking a break [3]. The Virtual learning system presents classes through video lessons, discussion forums, virtual tests, and real-time video conversations. Moreover, despite the possible uses of online courses, the scale of pupils who quit the classes has been considerably high. The current analysis also proves that the finalization scale in online courses is lower than the amount of those admitted to the classes [4]. According to the statistical consequences of classes conducted by California University on the Coursera podium, just 7% of students finished the classes. Hence, the divination of student dropouts in online courses is significant [3].

The dropout divination of online course methods exhibits a favorable possibility of overturning the formidable dropout of pupils at the starting stage and emerging maintenance scales. Based on these predictions, the teaching staff could take medication to manage the encouragement of learning students and stop them from dropping out of the classes [2]. Recently, there are 2 major research techniques for online course quitting divination: Conventional machine learning (ML) and deep learning (DL) models. Conventional ML research mainly utilizes ML-based categorization algorithms that need various guide functions to extract attributes which is an essential loss of time as well as human assets. Moreover, dropout prediction based on ML relies on the standard of the instruction specimens [5]. The models of DL have accomplished greater divination outcomes on online close dropout datasets when compared with traditional ML methodology [2]. Some research utilized ML as well as DL for feature extraction purposes for student dropout detection.

Moreover, educational data mining (EDM) is the implementation of DM methods to intelligent data for solutions to education sector issues. EDM captivates the utilization of stats, conception, and ML methods for the examination as well as the estimation of knowledge data. Several EDM implementations involve the articulation of online learning schemes, congregating knowledgeable data, and creating divination of student performance [4].

The foremost intention of this research is to design and develop the fractional-iterative deep BiLSTM (F-itr Deep BiLSTM) model for predicting student dropout, in which the student's performance in the online portal will be considered in every online course. For this prediction, the data is collected from the dataset [6], which is forwarded to the encoder layer for embedding to encode the data. Moreover, the informative features are extracted from the input data and then converted as a feature vector, which is forwarded to the decoder layer. In the decoder layer, the developed F-itr Deep BiLSTM classifier is engaged in predicting dropouts based on behavioral changes. Thus, the key significance of the developed research is mentioned below,

1.1. Fractional-Iterative BiLSTM

The non-locality of fractional operators and their flexibility is utilized in the iterative BiLSTM classifier for analyzing the information and predicting the student dropout in an efficient manner, and the fractional derivatives are also utilized for the optimal control problem while predicting the student dropout.

The organized format of the developed research manuscript is listed: the framework of the student's dropout from online courses is mentioned in section 1, and the review related to the development is briefly explained in section.2. Section 3 explains the developed F-itr Deep BiLSTM classifier for student dropout divination from virtual courses. The outcome of the developed model and the inference of the whole research are depicted in sections 4 and 5, respectively.

2. Literature Review and Challenges

The review of the related works for student dropout prediction is briefly explained in the below section, including their merits as well as demerits.

A deep analytical model was developed, as well as described in [1], to identify the dropout of students from online courses to spontaneously extract attributes from the real data and divination if every pupil will quit or finish classes. The described model utilized dual DL methods, Convolutional neural network (CNN), and Long-short-term memory (LSTM), which improved the performance compared with other models. However, it does not consider the student's performance. In [2], a novel DL model is initiated to foresee the students quitting as per learner behavior data. The CNN method is employed in the feature extraction mechanism to obtain informative features. Due to this, it attained high accuracy. The main disadvantage of the developed model is a lot of training data; moreover, they fail to encode the position and orientation of objects. A novel Gated Recurrent Units and autoencoders approach was designed and presented in [3] with the purpose of exploiting the secret space details and time-related data from disciple orientations to resolve the issue regarding the quitting divination of students. The developed research was performed well in complex cases with the best accuracy; however, it was not well-suitable for prediction where the input data were not a sequence. Besides, the computational complexity of the developed model is high. A model, which is depicted in [4] for the purpose of predicting the pupils quitting from online classes, provided a pair of attributes manufactured from students' day-to-day improvement of learning. The prediction and evaluation of the presented phase were performed with the function of an ML-based random forest (RF) classifier. The developed model predicts the quitting or prolongation of students during every given period in the online course. The developed model produced good, easily understood prediction results and efficiently handle large datasets, yet it was subject to interference and noise. The contributions of ML-based classifiers rely greatly on the standard of instructed representatives. In order to overcome this issue, a novel model was developed based on the weight calculation RF phase. As the conventional region method does not influence the address of the representative, then a new definition of neighborhood, which is referred to as the max region, is given at first, which is not only connected to the gap among representatives but also connected to the address of the representatives. The bio-inspired particle swarm optimization was well-tuned with the weight of the ML-based RF classifier, which increases the convergence speed, but prediction accuracy is very low.

2.1. Challenges

- In the baseline methods, feature extraction is achieved through feature manufacturing, which is constrained by the person's time limitations as well as observation since it takes a long time to extract the features from raw data [1].
- The conventional community description is only connected to the interval among representatives and not connected to the address details of the representatives. The instruction representatives in ML are addressed. Thus, the definition of the conventional region is inappropriate for direct utilization in ML applications [5].
- The selection and detection of essential attributes are a few complexities for scholars due to the variation in a podium that involves online courses. Because of these feature selection challenges, the accuracy of the prediction for dropout students decreases [4].

3. Fractional-Iterative deep BiLSTM Model for Student Dropout Prediction

The foremost aspect of the developed research is designing and creating an F-itr Deep BiLSTM model for predicting student dropout, for which the pupils' contribution to the online portal in any online classes will be considered. At first, the data is gathered from the utilized dataset. Then for the purpose of prediction, the gathered data is forwarded to the encoder layer, which is embedded and encodes the input data suitable for the prediction. The encoder layer speeds up the data entry as well as increases the accuracy of data entry.

Moreover, it requires less storage space. In other words, the Statistical based behavioral features are learned, and the informative features are extracted to establish the feature for the prediction. In addition, feature extraction provides the benefit of decreasing the number of unnecessary data from the dataset. Once the feature maps are established, then the prediction is performed for informative features from the feature vector. Finally, the feature vector enters the decoding layer, where the BiLSTM classifier is employed in predicting the dropouts based on behavioral changes. The fractional operator's non-locality and flexibility are utilized in this research, and the optimal control problem is also reduced through the fractional derivatives. Figure 1 depicts the block diagram representation of the developed phase.

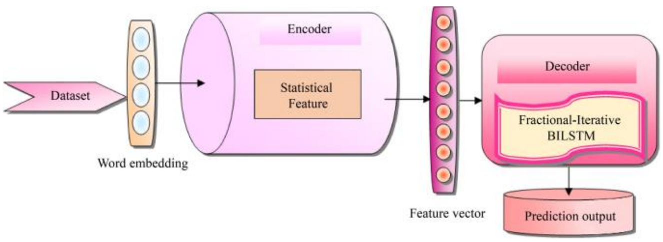


Fig. 1 Block diagram for the developed dropout student prediction

3.1. Encoder for Student Dropout Prediction

This proposed approach uses an encoder layer to increase the contribution of the successive learning consistent with the architecture. Moreover, it is used to extract the informative features based on statistics to create a feature map. Initially, the data is fed to the encoder layer, which encodes and embeds the input for prediction. The main task of the encoder layer in this developed phase is to understand a condensed and denoised presentation of the attributes to alleviate the scarcity occurrence.

3.1.1 Statistical Features

Mean

The mean for the student dropout prediction is estimated by the summation of the data point integrity to the total signal data points, which is expressed arithmetically as

$$m = \frac{1}{N}\sum_{l=1}^{N} h_1 \tag{1}$$

Here m is referred to as mean, Nsignifies the total data point, and $h_1$is referred to as the integrity data point.

Median

The median is considered the midway numerical dividing the greater half from the bottom half, which is utilized during the time of dataset outliners alter the average as the median hardly gets impacted by the outliers and is represented arithmetically as

$$M_n = \frac{\left(\frac{x}{2}\right)^{th} + \left(\frac{x}{2}+1\right)^{th} observation}{2} \tag{2}$$

where M_n is signified as the median, $x$signifies the amount of the greater half.

Standard Deviation

The term standard deviation is the evaluation of the angular of deviations that calculates the diffusion of an assemblage of data from its mean, which is arithmetically represented as,

$$\sigma = \sqrt{\frac{1}{N}\sum_{l=1}^{N}(f_l - m)^2} \tag{3}$$

where σ is signified as the standard deviation and $m$is referred to as the mean.

Variance

The Variance in the statistical features of the student dropout prediction is the estimation of the extent of data spreading out from the mean and is expressed arithmetically as

$$\sigma^2 = \frac{\sum_{z=1}^{Z}(x(z)-m)^2}{Z} \tag{4}$$

Maximum

The term maximum in the statistical features, otherwise called the greater restrict of an outlier, is the maximum integrity in an assemblage of integrity, prohibiting every outlier that is automated by utilizing the inter-quartile regulation or a statistical limitation of end user, which is expressed arithmetically as

$$l = P(0.25) - k[P(0.75) - P(0.25)] \tag{5}$$

where l denotes the statistical maximum.

Minimum

The term minimum is referred to as the statistical features, which is otherwise termed as the less restriction of outliers that is considered as the minimum integrity in an assemblage, dropping out every outlier, and automated by utilizing the inter-quartile regulation or a statistical limitation of the end user, which is expressed arithmetically as

$$h = P(0.75) + K[P(0.75) - P(0.25)] \tag{6}$$

where h represents the statistical minimum.

Skew

The term skewness in the statistical features is contemplated as a torsion that transmits the arranged curve of the bell available in the DTS as well as calculates the parallelism in the data by measuring if the input data is calculated, and is presented mathematically as

$$skewness = \frac{m_3}{\sigma^3} \tag{7}$$

where, $m_3$is signified as the 3rd moment of the mean-variance σ.

Kurtosis

The kurtosis measures the scale of disinclined data in the distribution of data, which is an estimation to discover if the input data is a system organized in a normal distribution and is expressed arithmetically as

$$kurtosis = \frac{M_4[x(n)]}{M_2^2[x(n)]} \tag{8}$$

where, M_4 is defined as the 4th moment of $x(n)$ and M_2^2 is referred to as the 2nd moment of $x(n)$, and with the normal distribution k=3, then the formula is

$$Sd_k = M\frac{M_4[x(n)]}{M_2^2[x(n)]} - 3 \tag{9}$$

here, $Sd_{kurtosis}$ is signified as the normal distribution of kurtosis.

3.2. Decoder for Student Dropout Prediction

Once trained, the encoder layer provides a feature vector for input, which the decoder uses to construct the input with the features that matter the most to make the reconstructed input recognizable as the actual input. In this developed research, the decoder layer utilized the F-itr Deep BiLSTM classifier for the purpose of student dropout prediction.

3.2.1. Fractional-Iterative BiLSTM Classifier Model for Prediction of Student Dropout

Egret is the combined term of tetrad breeds of birds like the Yellow-billed Egret, the Little Egret, the Great Egret, and the Middle Egret, which are popular for their splendid white feathers. Most of the Ardea Alba populate coastal islands, estuaries, coasts, lakes, rivers, rice paddies, streams, ponds, and sloughs close to their coasts. The Ardea alba is normally monitored in sets or else in mini crowds. Nevertheless, an extensive covey of more than or nearly hundreds could be discovered. Moreover, the Great Ardea albas' with a combative exploration scheme provides the right amount of large consumption of energy for the possible significant comebacks, while the Snowy Ardea albas' with a Stand by and watch approach, provides the right amount of less energy expenditure for little but huge dependable gains.

Step 1: Initialization

An LSTM layer consists of several cells which hold weights and biases corresponding to gates. A data sequence is inputted to the first cell, time-step by time-step. During this process, hidden states are created at each step of the sequence, which get passed back into the cell, and are expressed as

$$Solution, S = \{W_{fh}, W_{fx}, W_{ih}, W_{ix}, b_f, b_i, W_{ch}, W_{cx}, b_c, W_{oh}, W_{ox}, b_o\}$$

Step 2: Fitness Evaluation

The best solution is decided based on fitness, and here the accuracy is assumed as fitness. The solution corresponding to the maximum value of accuracy is selected for the classifier learning.

Step 3: Solution Updation

An iterative algorithm based on optimization ensures an improved performance

Global Aggression Stage

The solution update is based on the current solution and memory. Let us derive the solution update in the global aggression stage. Normally, the aggressive updation is mathematically modeled as,

$$S^{th} = S^t + exp(\frac{-t}{0.1t_{max_{1_{2_g}}}}^t \tag{10}$$

where, $D$is referred to as the distance between upper and lower convergence, and t is referred to as the iteration. Thus, the fractional character is imputed in expression (10).

$$S^{th} = S^t - a_1a_2S^t + exo(\frac{-t}{0-1t_{max_{1_{2_g}}}} \tag{11}$$

$$S^{th} - S^t = -a_1a_2S^t + exp(\frac{-t}{0.1t_{max_{1_{2_g}}}} \tag{12}$$

$$S^{th} = S^t(1 - a_1a_2) - \frac{1}{2S^{t-1}} - \frac{1}{6(1-b)S^{t-2}} - \frac{1}{24b(1-b)(2-b)S^{t-3}} + exp(\frac{-t}{0.1t_{max_{1_{2_g}}}}$$

where, S_g is referred to as the global solution, which is estimated using the previous convergence rates.

Local Aggressive Strategy

The multimodal issues preserve various local optimal solutions and numerous heights to obstruct the analysis of the local solution.

$$S^{th} = \left(1 - r_i - r_g\right)S^t + r_h.S_{personalbest} + r_g.S_g \tag{13}$$

$$S^{th} - S^t = \left(1 - r_i - r_g\right)S^t + r_h.S_{personalbest} + r_gS_g \tag{14}$$

$$D(S^{t+1} - S^t) = S^t\left(-r_i - r_g\right) + r_h.S_{personalbest} + r_gS_g \tag{15}$$

$$S^{th} = S^t - \frac{1}{2S^{t-1}} + S^t\left(-r_i - r_g\right) + r_h.S_{personalbest} + r_g.S_g \tag{16}$$

$$S^{th} = S^t\left(1 - r_i - r_g\right) - \frac{1}{2S^{t-1}} + r_h.S_{personalbest} + r_gS_g \tag{17}$$

This aggressive strategy contains the ability to obtain the global solution by terminating the disturbances from various local optimum points.

$$MSE = \sqrt{\frac{1}{z}\sum_{i=1}^{z}(o_i - p_i)^2}$$

Step 4: Re-evaluation of Fitness and Declare the Global Best Solution

The Egret optimization is sufficient for all sorts of controls and has the ability to fast convergence in simplistic issues when managing the splendid abstraction and hardiness for difficult issues. The aggressive strategy makes certain that the algorithm is exceptionally investigational and would not possibly fill into the local optimum, showing a splendid result in terms of analysis and manipulation.

Step 5: Terminate

Repeat the steps for the maximal iteration and declare the solution for updating the deep BiLSTM model.

4. Result and Discussion

The Student dropout prediction is established by utilizing the datasets [8] and [9] and the iterative BiLSTM classifier, and the achievement of the method is presented in this sector.

4.1. Experimental Setup

The developed BiLSTM classifier method for the classification of Land cover change is applied in Python. The arrangement involved in this scheme is defined as Python 3.7.6 assembled in Pycharm 2020 and executed in Windows 10 system software**.**

4.2. Dataset Description

The datasets utilized in this investigation are collected from the KDD Cup 2015 dataset (DTS) [6], which is based on Massive Open Online Course (MOOC) dropout prediction.

4.3. Performance Metrics

4.3.1. Accuracy

The student dropout prediction accuracy experiment is estimated according to the probability of the right prediction to the total number of predictions achieved from the prediction dataset by utilizing the below expression

$$AC = \frac{C_p}{T_p}$$

where C_p is defined as the correct prediction, $AC$is referred to as accuracy, and $T_p$is defined as the total prediction.

MSE

The occurrence of an error during the period of Student dropout prediction is known as Mean Square Error (MSE) as well as expressed arithmetically as

where z is defined as the monitoring size, whereas $o_i$indicates the real value, and p_i represents the predicted value.

Sensitivity

Sensitivity is the measurement of positive case predicted by the developed methods with the collected prediction datasets by utilizing the following expression,

$$SN = \frac{T_p}{T_p + F_p}$$

here, T_p is indicated as the true positive,F_p is indicated as the false negative and $SN$is indicated as sensitivity.

Specificity

The specificity of the prediction method is measured according to the probability of true negatives to the total negatives available in the dataset by utilizing the following expression.

$$S_p = \frac{T_n}{T_n + F_n}$$

where T_n is signified as a true negative, S_p is defined as specificity, and F_n is signified as a false positive.

4.4. Performance Analysis

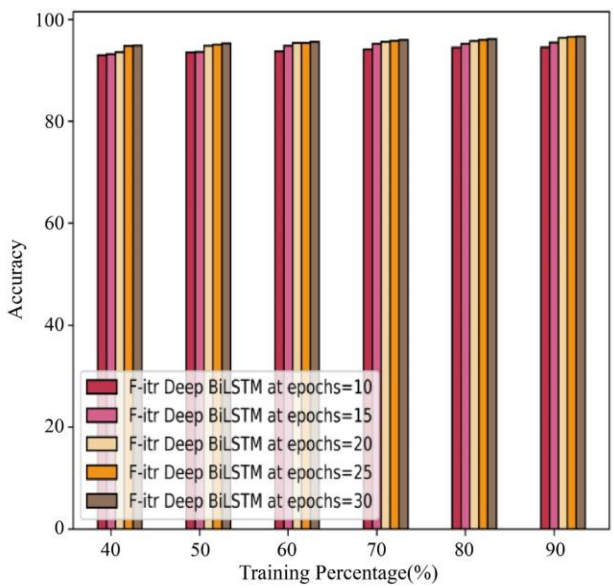
This sector explains the performance of the F-itr Deep BiLSTM classifier in terms of their performance metrics based on the DTS [6].

4.4.1. Performance Analysis based on DTS

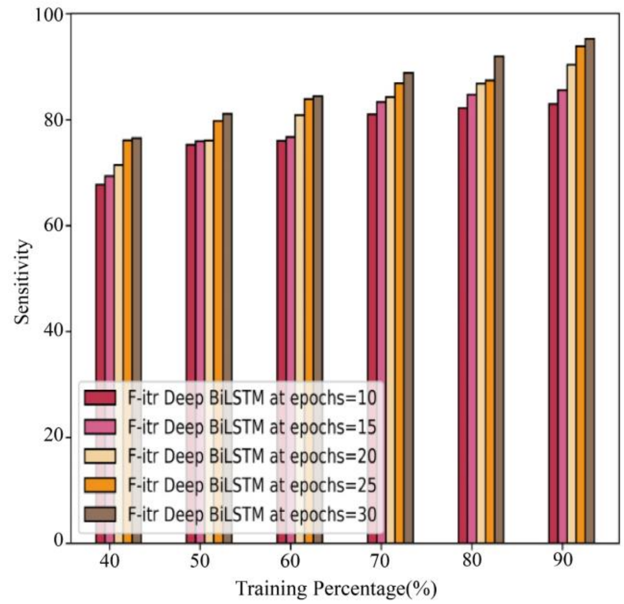
The performance examination of the iterative BiLSTM classifier for student dropout prediction as per the training point is presented in Figure 2 in terms of the performance metrics of the developed classifier, like accuracy, MSE, sensitivity, as well as specificity. The accuracy rate of the classifier at training point 90 with Epoch 10 is valued at 94.53%, with Epoch 15 is 95.49%, with Epoch 20 is 96.37%, with Epoch 25 is 96.58%, and with Epoch 30 is 96.71%. Whereas the MSE value at training point 90 with Epoch 10 is 0.13%, with Epoch 15 is 0.12%, with the Epoch value 20 is 0.12%, with Epoch value 25 is 0.11%, and with Epoch value 30 is 0.11% as well as the sensitivity value at training point 90 with Epoch 10 is 82.96%, with Epoch 15 is 85.60%, with Epoch 20 is 90.37%, with Epoch 25 is 93.87%, and with Epoch 30 is 95.31%. The specificity value at training point 90 with Epoch 10 is valued as 95.07%, with Epoch 15 is 95.57%, with Epoch 20 is 95.71%, with Epoch 25 is 96.48%, and with Epoch 30 is 97.01%.

4.5. Comparative Analysis

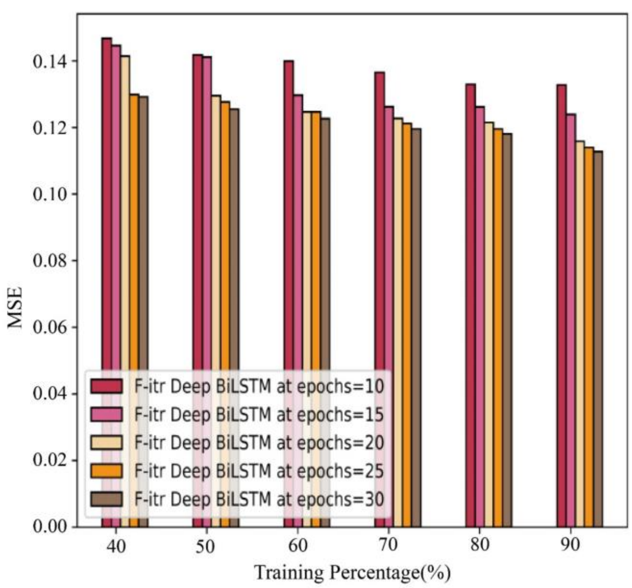
The comparative evaluation of the developed method is estimated with several previous methods like Support Vector Machine (SVM) (MTD1) [7], multilayer perception (MLP) classifier (MTD2)[8], light gradient-boosting machine (Light GBM) (MTD3) [9], Decision tree classifier (MTD3)[10], and Deep Neural Network (MTD4)[11].



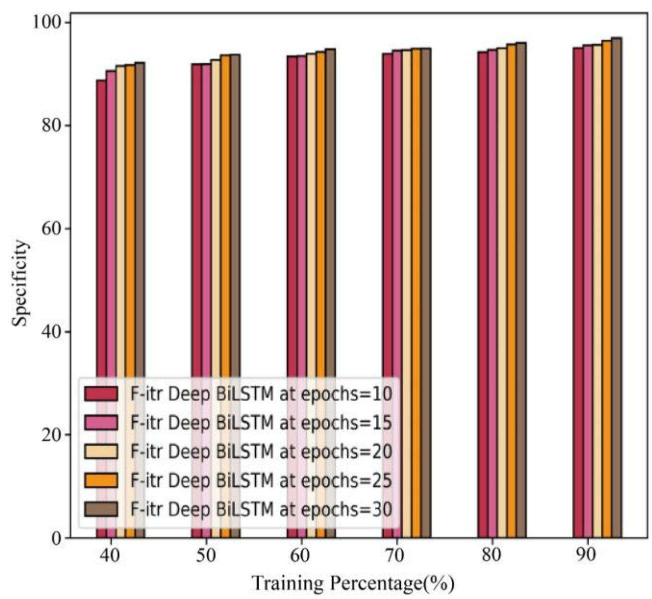
a) Accuracy



c) Sensitivity



b) MSE



d) Specificity

Fig. 2 Performance analysis of the iterative BiLSTM model using DST a) accuracy, b)MSE, c) Sensitivity and d) Specificity

4.5.1. Comparative Analysis based on DTS

The comparative examination is estimated per the DST [6] and is achieved based on accuracy, MSE sensitivity, and specificity, as represented in Figure 3. The accuracy rate of the proposed method at training point 90 with epoch is valued as 96.44%, which is increased by 13.82% to the existing MTD1, 6.54% to MTD2, 4.68% to MTD3, 2.11% to MTD4, and 1.69% than MTD5. Whereas the sensitivity rate is valued as 95.22%, which is an increase of 60.06% from the existing MTD1, 29.45% from MTD2, 15.33% from MTD3, 12.13% from MTD4, 9.58% from MTD5 and the specificity rate is valued at 96.88% which is an increase of 9.18% than the existing MTD1, 5.93% than MTD2, 5.66% than MTD3, 4.26% than MTD4, and 4.03% than MTD5 respectively, as well as the MSE rate at training point 90 is valued 0.09% which is decreased by 10.75 than MTD1, 10.59% than MTD2, 9.23% than MTD3, 7.60% than MTD4, and 6.93% than MTD5. Figure.2 presents the Comparative Evaluation of the proposed method using DST1.

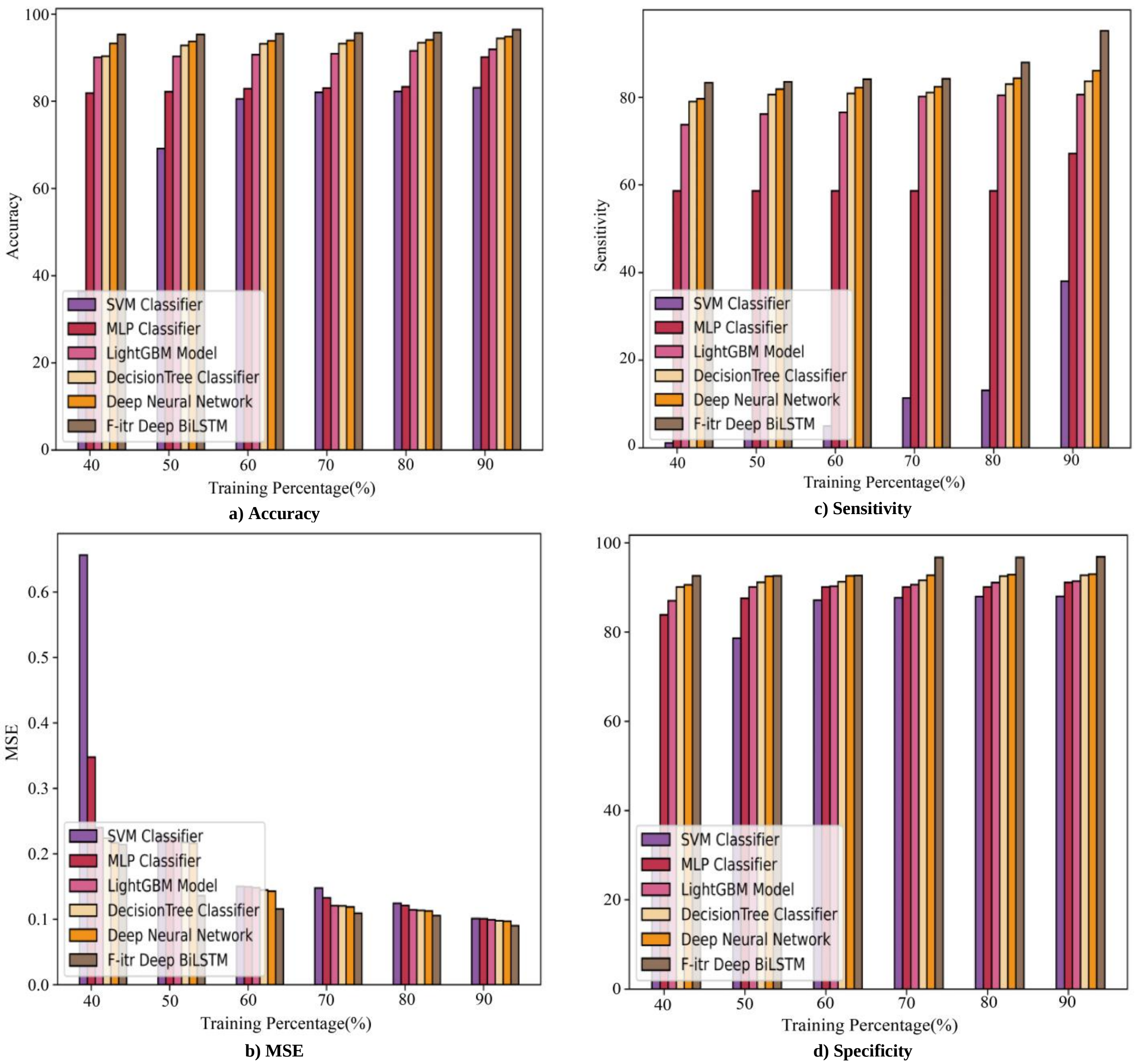


Fig. 3 Comparative analysis using DST a) accuracy b) MSE c) Sensitivity and d) Specificity

5. Conclusion

Student dropout prediction from online courses using the Fractional-Iterative BiLSTM classifier is performed in this research. The prediction of the students quitting will benefit educational institutions in estimating and understanding the knowledge-gaining methods of students across the varied interchange of the students. This would also allow educational institutions to establish techniques to encourage and produce learner improvement. In this research, a Fractional-Iterative BiLSTM classifier is developed to predict student dropout accurately. The encoder layer is used to extract informative and behavioral features based on Statistical features. The decoder layer utilized the developed Fractional-Iterative deep model, which effectively classifies the feature vector for the purpose of predicting student dropout. The accomplishment of the research is evaluated by estimating the enhancement, and the developed model obtained an increase of 96.71%% for accuracy, 95.31% for sensitivity, as well as 97.01%.for specificity, which shows that the efficiency of the method, and the MSE is reduced by 0.11%. In the future, the research can be improved by utilizing hybrid optimization. Therefore the effectiveness could be further enhanced.

References

[1] Ahmed A. Mubarak et al., “Deep Analytic Model for Student Dropout Prediction in Massive Open Online Courses,” *Computers & Electrical Engineering*, vol. 93, p. 107271, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Qian Fu et al., “CLSA: A Novel Deep Learning Model for MOOC Dropout Prediction,” *Computers & Electrical Engineering*, vol. 94, p. 107315, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Bardh Prenkaj et al., “Hidden Space Deep Sequential Risk Prediction on Student Trajectories,” *Future Generation Computer Systems*, vol. 125, pp. 532-543, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Sheran Dass, Kevin Gary, and James Cunningham, “Predicting Student Dropout in Self-Paced MOOC Course Using Random Forest Model,” *Information*, vol. 12, no. 11, p. 476, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Cong Jin, “Dropout Prediction Model in MOOC Based on Click Stream Data and Student Sample Weight,” *Soft Computing*, vol. 25, pp. 8971-8988, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] The KDD Cup 2015 Dataset, 2023. [Online]. Available: https://data-mining.philippe-fournier-viger.com/the-kddcup-2015-dataset-download-link/

[7] Liming Liu et al., "An Improved Nonparallel Support Vector Machine," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 5129-5143, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Meha Desai, and Manan Shah, "An Anatomization on Breast Cancer Detection and Diagnosis Employing Multi-Layer Perceptron Neural Network (MLP) and Convolutional Neural Network (CNN)," *Clinical eHealth*, vol. 4, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Altyeb Altaher Taha, and Sharaf Jameel Malebar, "An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine," *IEEE Access*, vol. 8, pp. 25579-25587, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Fei Wang et al., "A Linear Multivariate Binary Decision Tree Classifier Based on K-Means Splitting," *Pattern Recognition*, vol. 107, p. 107521, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Wojciech Samek et al., "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications," *Proceedings of the IEEE*, vol. 109, no. 3, pp. 247-278, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] W. Xiong, D. Litman, and C. Schunn, "Designing and Evaluating Automated Feedback for Improving Revision of Argumentative Writing," *International Journal of Artificial Intelligence in Education*, vol. 28, no. 2, pp. 152-189, 2018.

[13] M. C. Shanker, and M. Vadivel, "Hybrid Transfer Learning of Mammogram Images for Screening of Micro-Calcifications," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 8, pp. 40-47, 2022. [CrossRef] [Publisher Link]

[14] Hassan, S., and Elgendi, M., "Predicting Student Dropout using LSTM," *Procedia Computer Science*, vol. 159, pp. 200-209, 2019.

[15] Jiazhen He et al., "Identifying At-Risk Students in Massive Open Online Courses," *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, pp. 1749-1755, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[16] I. Ganguly, and S. Bandyopadhyay, "Student Dropout Prediction in MOOCs Using Popular Deep Learning Models: A Survey," *IEEE Access*, vol. 8, pp. 20830-20841, 2020.

[17] S. Boyer, and K. Veeramachaneni, "Transfer Learning for Predicting Outcomes of Newly Deployed MOOCs," *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 217-225, 2017.

[18] S. Tang et al., "Predicting MOOC Performance with Week 1 Behavior," *Education and Information Technologies*, vol. 24, no. 4, pp. 2695-2706, 2019. [Google Scholar] [Publisher Link]

[19] Neha Chaudhary, and Priti Dimri, "Enhancing the Latent Fingerprint Segmentation Accuracy Using Hybrid Techniques – WCO and BiLSTM," *International Journal of Engineering Trends and Technology*, vol. 69, no. 11, pp. 161-169, 2021. [CrossRef] [Publisher Link]

[20] Z. Zhou et al., "DeepMooc: A Deep Learning Framework for MOOC Dropout Prediction," *IEEE Access*, vol. 8, pp. 123262-123275, 2020.

[21] D. Rajendran, and M. Dinakaran, "Deep LSTM Approach for Dropout Prediction of Students in Distance Education," *Education and Information Technologies*, vol. 25, no. 2, pp. 1167-1182, 2020.

[22] S. Jiang et al., "Predicting MOOC Performance with Social Network Analysis," *Proceedings of the 19th International Conference on Artificial Intelligence in Education*, pp. 229-238, 2018.

[23] Justin Reich, and José A. Ruipérez-Valiente, "The MOOC Pivot," *Science*, vol. 363, no. 6423, pp. 130-131, 2019. [CrossRef] [Publisher Link]

[24] Scott Crossley et al., "Predicting Success in Massive Open Online Courses (MOOCs) Using Cohesion Network Analysis," *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, pp. 53-57, 2018. [CrossRef] [Publisher Link]

[25] M Mamatha et al., "Enhanced Sentiment Classification for Dual Sentiment Analysis using BiLSTM and Convolution Neural Network Classifier," *International Journal of Engineering Trends and Technology*, vol. 70, no. 3, pp. 151-161, 2022. [CrossRef] [Publisher Link]

[26] Dragan Gašević et al., "Where is Research on Massive Open Online Courses Headed? A Data Analysis of the MOOC Research Initiative," *International Review of Research in Open and Distributed Learning*, vol. 15, no. 5, pp. 134-176, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[27] J. Jia et al., "MOOC Dropout Prediction: Lessons Learned from Making Feature Engineering Transparent," *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education*, pp. 134-140, 2021.

[28] J. Taylor, J. Silverman, and R. Shoop, "Ensemble Machine Learning MOOC Dropout Prediction Using Course-Level Data," *Proceedings of the 11th International Conference on Computer Supported Education*, pp. 83-92, 2019.

[29] Y. Zhang et al., "Co-Training for Dropout Prediction in MOOCs," *International Journal of Artificial Intelligence in Education*, vol. 30, no. 2, pp. 192-220, 2020.

[30] Jacob Whitehill et al., "Beyond Prediction: First Steps toward Automatic Intervention in MOOC Student Stopout," *Proceedings of the 8th International Conference on Educational Data Mining*, pp. 171-178, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[31] Arti Ramesh et al., "Understanding MOOC Discussion Forums using Seeded LDA," *Proceedings of the 9th International Conference on Educational Data Mining*, pp. 126-133, 2014. [Google Scholar] [Publisher Link]

[32] C. Chen et al., "Predicting Student Attention in MOOCs: A Deep Learning Approach Using Multimodal Video Features," *Proceedings of the 10th International Conference on Learning Analytics & Knowledge*, pp. 276-285, 2020.

[33] Heng Luo, Anthony Robinson, and Jae-Young Park, "Peer Grading in a MOOC: Reliability, Validity, and Perceived Effects," *Journal of Interactive Online Learning*, vol. 15, no. 1, pp. 25-38, 2017.