Prediction of Student's Interest on Sports for Classification using Bi-Directional Long Short Term Memory Model

A. Basheer Ahamed[1], M. Mohamed Surputheen[2]

abammb2@gmail.com, msurfudeen@yahoo.com

[1]Research Scholar, [2]Associate Professor
PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous)
(Affiliated to Bharathidasan University), Tiruchirappalli, Tamil Nadu, India.

Abstract

Recently, parents and teachers consider physical education as a minor subject for students in elementary and secondary schools. Physical education performance has become increasingly significant as parents and schools pay more attention to physical schooling. The sports mining with distribution analysis model considers different factors, including the games, comments, conversations, and connection made on numerous sports interests. Using different machine learning/deep learning approach, children's athletic and academic interests can be tracked over the course of their academic lives. There have been a number of studies that have focused on predicting the success of students in higher education. Sports interest prediction research at the secondary level is uncommon, but the secondary level is often used as a benchmark to describe students' educational development at higher levels. An Automated Student Interest Prediction on Sports Mining using DL Based Bi-directional Long Short-Term Memory model (BiLSTM) is presented in this article. Pre-processing of data, interest classification, and parameter tweaking are all the essential operations of the proposed model. Initially, data augmentation is used to expand the dataset's size. Secondly, a BiLSTM model is used to predict and classify user interests. Adagrad optimizer is employed for hyperparameter optimization. In order to test the model's performance, a dataset is used and the results are analysed using precision, recall, accuracy and F-measure. The proposed model achieved 95% accuracy on 400^{th} instances, where the existing techniques achieved 93.20% accuracy for the same. The proposed model achieved 95% of accuracy and precision for 60%-40% data, where the existing models achieved 93% for accuracy and precision.

Keywords:- Sports Mining; Education Performance; Interest Prediction; Bi-Directional Long Short Term Memory; Physical Education.

1. Introduction

Factors like active involvement in learning, educational experiences, collaborations and communication are combined to determine whether students are satisfied with their educational experience, and a deeper understanding of these elements and how they interact is critical for many higher education institutions [1]. Many new professional women's sports leagues (such as basketball, soccer, and fast-pitch softball) have formed in recent years, indicating an increased demand for female-dominated sporting events [2]. More women are playing intercollegiate sports today because of Title IX legislation and the growing popularity of women's professional sports leagues [3]. The number of women's sports teams and athletes in the National Collegiate Athletic Association (NCAA) has increased dramatically since Title IX was passed. Despite the fact that men's sports have historically been the cash generators, many Division I institutions still have far more men's sports than women's sports [4].A country's sporting talent is primarily discovered through its educational institutions. For the advancement of sports in any country, students' passions are centred on various sports. Every organization entertains a sizable number of students throughout their education, each with a unique set of interests and a willingness to take part in a variety of events [5]. Identifying the pupils who are most interested in a particular sport will aid in determining the student's level of interest in sports [6]. In a nation with such a large population, there are talents everywhere, so it is crucial to utilize the students with diverse sporting abilities for the development of sports in India.Higher education is typically viewed as a service sector due to its emphasis on offering students high-quality services in an environment that is becoming more competitive [7]. In this situation, it becomes increasingly crucial to meet students' needs and expectations in order to recruit and retain good students. There are two methods for rating the educational experience [8]. The first approach is used to assess teaching and learning whereas the second approach is used to assess the entire student experience. Considering students as customers of education [9], their degree of gratification plays a significant influence on the

Manuscript received October 5, 2022
Manuscript revised October 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.10.32

success of the institution in terms of both effectiveness and recruitment. Student life is a network of interconnected events that affects student satisfaction, and student gratification is adopted as a short-term attitude as a result of an appraisal of students' educational experiences [10]. Having a satisfied student population has many advantages including greater satisfaction and commitment. Every higher education institution places a high value on keeping students on campus until they have finished their degrees. Customer satisfaction with the level of service is important [11].

Although it is a theoretical field, the classroom flip is rarely used in practical classes like physical education. It is now possible to combine the internet by means of physical education classes along with new teaching style [12]. Teaching athletic skills in a typical classroom is reimagined using a flipped classroom paradigm. In order to use this method of training, physical education instructors must produce short videos based on the content of their lessons. The length of a short film should be between five and ten minutes, so that the lesson's content may be fully explained [13]. The quality of video created by each differs, and so does the level of expertise.

As the internet grows and becomes more widely used, students are depending on it more and more. An educational revolution is inevitable when internet-based flipped classrooms are introduced. As a result of this transformation, learning and teaching have evolved. Student learning is more autonomous, and teachers are better able to tailor their instruction to each student's ability [14]. Individuality is an important goal for college students who desire to think for themselves. With the combination of educational learning quality, they make full use of their capacity to independently consult and develop a preliminary comprehension of subject. Hadoop's distributed storage scalability and parallel computing capabilities demand an effective framework for processing huge volume of big data, allowing users to concentrate on domain business analytic and uncovering value across a number of disciplines. University sports choice support systems can facilitate challenges and obstacles on the path to achieving sports growth [15], provide strategic support for sports development, and accelerate the development of physical education in a globalized and information-driven world.

In this research work, sports interest among students is identified using DL-LSTM model. Initially, pre-processing is carried out by data augmentation and the performance of the proposed model is tested by using parameters such as precision, recall, f-measure and accuracy. The related works for the proposed model is given in Section 2. The proposed methodology explanation is depicted in Section 3. The validation of the proposed model with existing techniques is given in Section 4. Finally, the conclusion of the research work is presented in Section 5.

2. Literature survey

Wei Xu et al. [16] examined the use of neural knowledge in flip classroom in order to increase sports skill acquisition for students. A strategy was followed to predict college athletes' performance that utilised a neural network powered by particle swarm optimization for high efficiency and accuracy.

In order to determine which professional field their preferences and aptitudes, Roy KS et al. [17] suggest that it was crucial for children to assess their abilities and defined their interests while they were still in school. This would help them perform better and stimulate their interest in pursuing a career in their field of choice. After thoroughly analysing each candidate, career recommender systems help recruiters for choose where to position new students based on their performance and other factors. Machine Learning (ML) technique succeeds thorough data collection. In this case, the algorithm considered both the input and the output as data. Numerous factors must be considered in order to accurately predict a student's academic performance in various areas and specialties. This study focused mostly on predicting the career path of computer science students.

A study by Haiyun Z et al. [18] focused on the classification and information aspects. It also explained how the Hadoop cloud computing platform was developed and utilized. Models of cloud-based data collection and prediction models were included in this article one-piece learning method. With its ability to handle data of varying sizes, Hadoop provided this paradigm. On the Hadoop platform, the training skill was evaluated in a number of scenarios to make sure it was accurate and of the best quality. The Spark framework for this research focuses on building computational engines and improved the prediction models by utilizing large data and Hadoop learning. Hence, the proposed model provided high-accuracy and timely genuine hurricane forecasting.

Sorkkila M, et al. [19] investigated about the co-developmental dynamic of sport using SEM. The proposed model served as a predictor to measure the dimension of sports activities. Cynicism and self-doubt in the same field were negatively linked to mastery goals in sport or school. School-related pessimism was also substantially connected with school-related performance goals. Using the findings, healthcare providers helped student athletes avoid burnout before it occurred.

Zhang X et.al [20] suggested a method for predicting the achievements of the sports students. The development of scientific training plans for athletes relies heavily on the ability to accurately predict their future performance. College students' group sports achievements helped to predict students' follow-up learning outcomes. Many schools' sports performances were evaluated and analysed in this study from a standpoint of strength, endurance and sensitivity. The sporting achievements of notable universities were also compared and evaluated. This paper analysed and examined the situation of each school in order to discover how they achieved excellent results in physical education and proposed teaching strategies to improve physical education results.

Ting J et.al [21] found that the interest in various aspects of sports such as physical education and developing sports skills were mediated by a number of variables. The results of two surveys were gathered. The initial poll included 339 Taiwanese undergraduates, 232 of whom were male and 107 of whom were female. A total of 233 males, 98 women, and 2 students who did not specify their gender took part in the second survey, for a final calculation of 333 participants. According to the findings, sports habits and compulsive internet use were moderated by an interest in physical education (IPE). For as long as the student maintained a regular exercise routine that boosted IPE, they would be less prone to Internet addiction. It is emphasized how important it was to ignite students' IPE.

In [22] Basheer Ahamedet.al suggested a Sports Distribution Analysis Model (SDAM) as a means of enhancing performance. The sports distribution analysis approach was used to identify the sports interests of various students by their behaviour in the academic and social networks. The logs of sports activities were divided into a number of time windows, and the approach estimated the frequency of sports activity for each interest of those time windows. The approach performed distribution analysis and measured the elements affecting sports distribution based on these data. The approach predicted interest based on these numbers. There was a lower false positive rate with the SDAM.

Ma B et al. [23] used mobile technology for the design and development of an athlete training procedure, and monitoring scheme that employed GPS to gather real-time location information of athletes. In this study, synchronized tracking analysis was utilized to examine an athlete's specialized sports-specific functions. The training plan preparation provided by the coach was based on the brain functions and conditions, nutrition management, etc.

Zhang R et.al [24] suggested the method using data from fitness tests. Teachers utilized Few-Shot Learning technology to extract essential facts from the data, allowing them to deliver more scientific and effective feedback from the students. Here, the design and implementation were utilized to evaluate students' ability to do physical tests using mobile edge computing, as well as findings. As a result of the improved readability and efficiency of the fitness assessment software, students now had an improved understanding of their physical fitness levels. Now, teachers provided more advice and recommendations based on students' physical traits using data mining.

3. Proposed Model

The basic workflow of the proposed model is given in Figure 1, where each step is discussed briefly in the following sub-sections.

3.1. Data Description

The National Collegiate Athletic Association (NCAA) of Basic Academic Skills Study (BASS) is a comprehensive scale meant to examine the interests, attitudes, and academic skills of student athletes. Three main components make up the BASS, one of the few substantial datasets on subscale was created to measure students' academic and social accomplishments and failures, as well as their personal goals and attitudes towards college. Students in high school and college can use the Social and Group Experiences (SAGE) subscale to assess certain aspects of their educational experiences. In order to gauge a student's current proficiency in reading, writing, math, and general knowledge, the Mini-Battery of Achievement (MBA) subscale was developed.

The Progress in College (PIC) and SAGE subscales were used in a secondary analysis for this study. A total of 410 freshmen from 21 Division I colleges and universities was surveyed in the year 1996–97. High-profile sports consisting of 25.1 % of the participants' time; 74.9% of those who took part were in low-profile sports; 54.6 % of those who took part in high-profile sports are male and 45.4% of those who took part in low-profile sports are female, and 75.5 % are white. Latino/Hispanic American, American Indian, and other ethnic groups were also included in the list of other ethnicities.

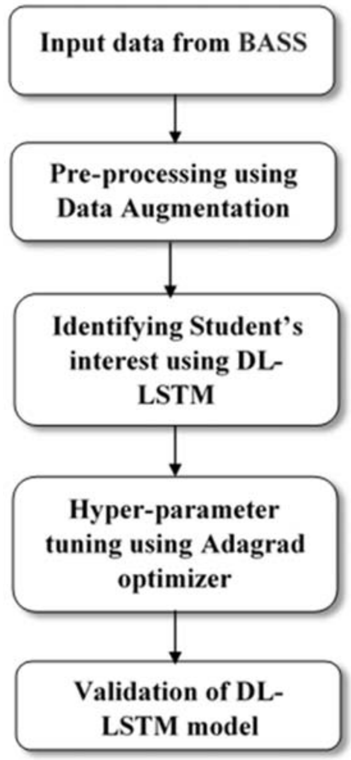

Figure 1: Workflow of proposed model

3.1.1. Variables of data

There are three kinds of variables used in this research work: (a) background factors (b) engagement variables (c) cognitive/effective outcomes. Among the variables,background factors were measured using gender, race/ethnicity, the primary field of study, and the profile level of the sport of the participants. In this study, interactions with faculty members and other students were all investigated as potential sources of student involvement.

It is possible to determine these four factors by taking the average across all items on each subscale. Each item was scored on a six-point Likert scale with a range of 1 to 6. (very often).Cultural attitudes and personal self-conceptions were employed as indices of student affective outcomes in this study. These two effective outcomes were assessed using the mean of each subscale's items, ranging from 1 (very strongly disagree) to 6 on a Likert scale of 1 to 6. (very strongly agree). The responses provided by the students showed growth in both their learning and communication abilities. For each of the four items, there were six possible Likert ratings, with 1 being the lowest and 6 being the highest.

3.2. Pre-processing

The noisy records in the dataset were removed during the pre-processing of sports data. To accomplish this, the whole dataset is read and used to identify the log's dimensions and attributes.

The procedure examines each piece of data and determines whether the traits or dimensions that need to be verified are present. Each data point's invalid values are either filled in or eliminated after a careful analysis of the existence and values of each dimension. The noisy records should be eliminated if any of the data points are discovered to be missing, it has been determined. On the pre-processed data points, categorization is done in the following phase.

A list of possible characteristics or dimensions is generated using the input data. As a result of this data, the entire dataset is free of spurious data points and acquired all the attributes.

3.3. Interest Prediction and Classification-DL-LSTM Model

The typical RNN structure is required for an LSTM. The Recurrent ANN problem of being unable to manage long-distance dependence is solved by using a variety of models for computing hidden states. However, despite the intrinsic virtues of the newly constructed model, these models do not learn the best potential of LSTM. The memory units in this LSTM technique have three gates each and act in a range of ways. When using the S vector as input, the pre-processed values are provided for certain conditions of the LSTM unit for input data for sports interest prediction of students. Figure 2 presents the basic architecture of the proposed LSTM model.

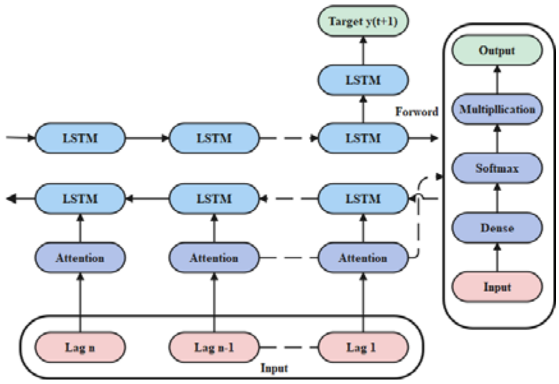

Figure 2: Basic Architecture of Proposed Model

The sigmoid function and the dot multiplication are defined in the following processing phrase.

A forget gate is implied by the f_t.:

$$f_t = \sigma(W_f w_t + U_f h_{t-1} + b_f) \quad (1)$$

The i_t signifies an input gate:

$$i_t = \sigma(W_i w_i + U_i h_{t-1} + b_i) \quad (2)$$

The $\tilde{c}_t$tanh defines the tangent hyperbolic function in the candidate memory cell condition recently consumed;

$$\tilde{c}_t = tanh(W_c w_t + U_i h_{t-1} + b_c) \quad (3)$$

The f_t and i_t series, which range from 0 to 1, are used to calculate the c t, or current time value, in memory cells. This means that new data is being stored in candidate unit $\tilde{c}_t$ as part of the processing of i_t. The purpose of $f_t \odot c_{t1}$ indicates that the data is conserved and it is unwanted for memories c_{t-1}.

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1} \quad (4)$$

The o_t indicates output gate:

$$o_t = \sigma(W_0 w_t + U_0 h_{t-1} + b_0) \quad (5)$$

h_t implies hidden layer state at time t:

$$h_t = o_t \odot tanh(c_t) \quad (6)$$

The LSTM only has a small amount of historical data. It can be used in a variety of contexts that will be available in the future. There are two layers in the bidirectional LSTM: a forward layer and a backward layer. The function can be found here: The forward layer uses data from a previous series, while the backward layer consumes data from a future series that has yet to be collected. These layers are then connected to form a single final layer. In this approach, all of the series context data is taken into account. At time $t-1$, the end result of the forward hidden unit is $\vec{h}_{t-1}$, and the simulated result of the backward hidden unit is $h^{\leftarrow}{}_{t+1}$, it is assumed that the time input is a word embedding w_t. Finally, the backward and hidden unit results at time t are identical:

$$\vec{h}_t = L(w_t, \vec{h}_{t-1}, c_{t-1}) \quad (7)$$

$$h^{\leftarrow}{}_t = L(w_t, h^{\leftarrow}{}_{t+1}, c_{t+1}) \quad (8)$$

The buried layer of the LSTM is referred to as L(.) Both the forward resultant vector and backward output vector must be combined in order to provide the text feature. H is shown as the number of cells in the concealed layer:

$$H_t = \vec{h}_t || h^{\leftarrow}{}_t \quad (9)$$

3.4. Parameter Tuning Process-AdaGrad Optimizer

This LSTM approach is used to slow down the rate at which new information is being absorbed during the training phase. Training a step size or "learning rate" improves the weight count. With a value between 0.0 and 1.0, the learning rate is a modelling hyper-parameter used to train a non-negative neural network (NN). These concerns can be solved by balancing the learning rate. Higher training epochs with modest weight changes are required to achieve minimum learning rates. The low training epochs and high learning rates are meant to produce large-scale changes. Simulating the rate of learning is a difficult task. A high learning rate leads to a diverging training process, while a low learning rate slows the process down. Various learning rates might be triggered throughout training to achieve an effective outcome.

Adagrad's parameters, which are used in gradient-based optimization, are susceptible to learning rates. Additionally, it produces both minor and major changes to factors related to salient traits and distinctive features. This makes it a good choice for handling sparse data. Stochastic gradient descent (SGD) is made more efficient with Adagrad, which is used to train massive neural nets. For sparse gradients, Adagrad optimizer is a gradient-based optimization model that works well in a gradient estimate, the learning rate replicates how much a parameter allows you to apply the opposite direction (g). It uses the qualities to determine the learning rate.

$$\theta_{t+1} = \theta_t - \frac{a}{\sqrt{\varepsilon + \Sigma \quad g_t^2}} \odot g_t \quad (10)$$

As represented in Eq. (10)), the fundamental function for parameter update is depicted as a function of t, a, the learning rate, g t, the gradient estimate.

4. Results and Discussion

The program of this research work is written in Python with the Keras platform and evaluation is conducted on Waikato Environment for Knowledge Analysis tool. The system with 12 GB DDR3 memory, a processor of 2.5 GHz and an Intel Core of i5 7200U is used for the proposed technique implementation.

4.1. Performances metrics

The performance of the classification algorithms was compared according to the criteria of accuracy, precision, and F-measure that is shown in Eq. (11-14).

$$Accuracy\ (AC) = \frac{A+B}{A+B+C+D} \quad (11)$$

$$Precision\ (P) = \frac{A}{A+B} \quad (12)$$

$$Sensitivity\ (S) = \frac{A}{A+D} \quad (13)$$

$$F - Measure\ (F-M) = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \quad (14)$$

Here, A represents the true positive, B describes the true negative, C presents the false positive and D depicts the false negative.

4.2. Performance Analysis of Proposed model

Here, there are two sets of data are considered for validation of DL-LSTM, which include 70% of training data-30% of testing data and 60% of training data-40% of testing data. Initially, the proposed model and existing models are tested without an Adagrad optimizer for both ratios of data in terms of various metrics, which is shown in Table 1. Figure 3 to 6 provides the graphical comparison of different ratio of training and testing data in terms of various metrics.

Table 1: Results of Deep learning models without optimization model

	Model	AC (%)	F-M (%)	S (%)	P (%)
70% Train-30% Test	RNN	86.23	83.50	86.20	82.80
	CNN	84.74	84.20	84.40	83.90
	LSTM	92.06	90.80	92.10	89.60
	Bi-LSTM	**96.49**	**96.50**	**96.50**	**96.50**
60% Train-40%Test	RNN	86.89	84.30	86.90	83.70
	CNN	84.93	84.50	84.90	84.20
	LSTM	92.12	90.90	92.10	89.70
	Bi-LSTM	**95.86**	**95.80**	**95.90**	**95.80**

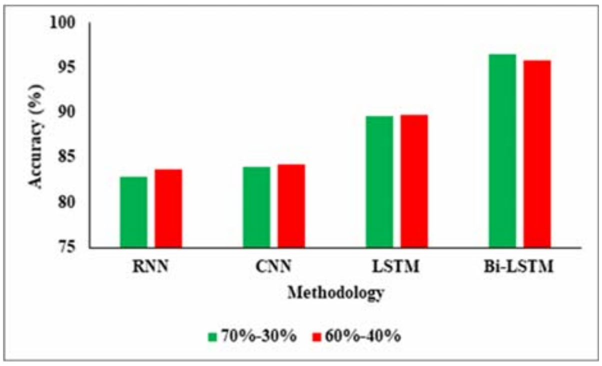

Figure 3: Accuracy comparison for existing with proposed technique

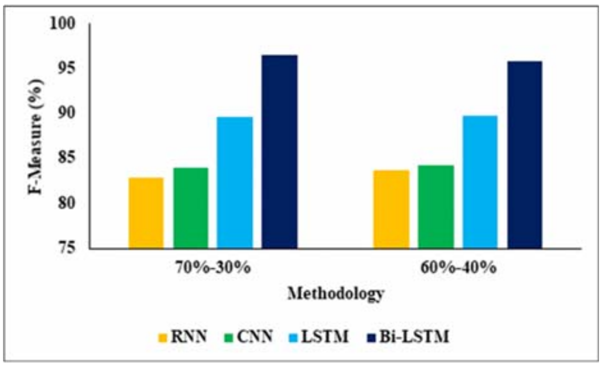

Figure 4: F-Measure comparison for existing with proposed technique

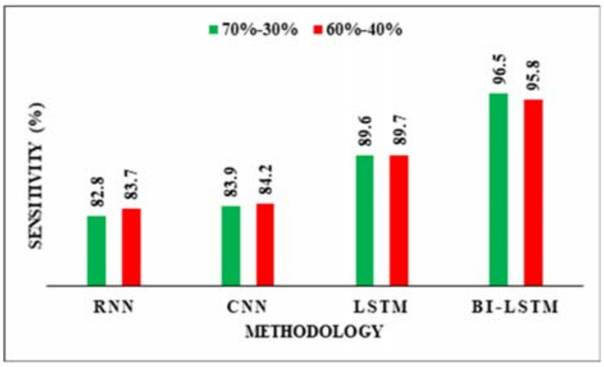

Figure 5: Sensitivity comparison for existing with proposed technique

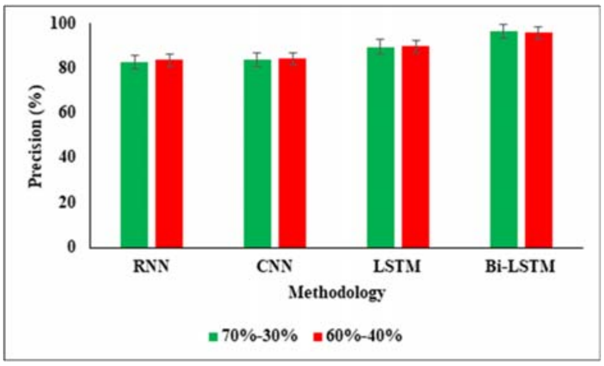

Figure 6: Precision comparison for existing with proposed technique

In this analysis, the Bi-LSTM model reaches the better accuracy of 96.49% for 70%-30% of data and also achieved the accuracy of 95.86% for 60%-40% of data respectively. LSTM achieved 92% of accuracy, RNN and CNN achieved nearly 85% of accuracy for 70%-30% of data. In the analysis of F-M, S and P, the existing techniques such as RNN and CNN achieved nearly 83% to 85%, LSTM model achieved nearly 90% to 92% and proposed model achieved 96% for the ratio of 70%-30% data. From this analysis, it is clearly proves that the proposed model achieved better performance than existing models. When the training data is 60% used and remaining 40% of data is used for testing process, at that time, the RNN and CNN achieved nearly 83% of P and F-M, 85% of S and accuracy. But the proposed model achieved 95% of AC, F-M, S and P. In this comparison analyses, the 70% Train- 30% Test data provide the better results in all classifier. When

Adagrad optimizer is implemented with proposed model, it achieved better performance, which is provided in Table 2. Figure 7 to 10 presents the graphical comparison of proposed model with optimizer that provides better results than Bi-LSTM without optimizer.

Table 2: Results of Deep learning models with optimization model

	Model	AC (%)	F-M (%)	S (%)	P (%)
70% Train-30% Test	**Bi-LSTM**	**97.02**	**97.00**	**96.73**	**97.28**
	LSTM	94.98	94.91	94.49	95.34
	CNN	93.92	93.83	93.38	94.30
	RNN	94.25	94.33	93.83	94.84
60% Train-40%Test	**Bi-LSTM**	**95.34**	**95.37**	**94.86**	**95.90**
	LSTM	93.16	93.27	92.79	93.76
	CNN	92.96	92.91	91.65	94.25
	RNN	94.79	94.78	90.08	95.53

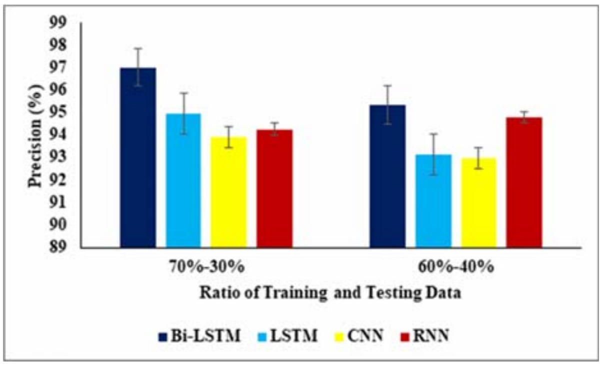

Figure 7: Precision comparison for DL techniques with optimizer

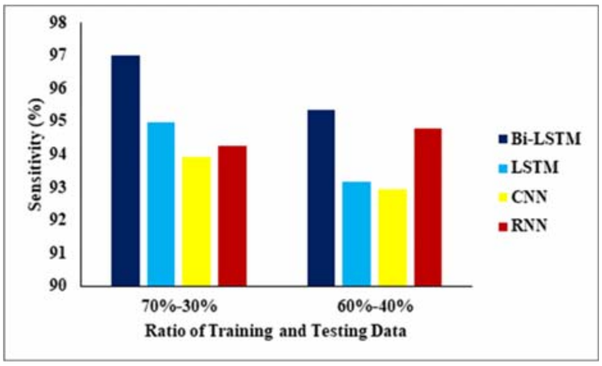

Figure 8: Sensitivity comparison for various techniques with optimizer

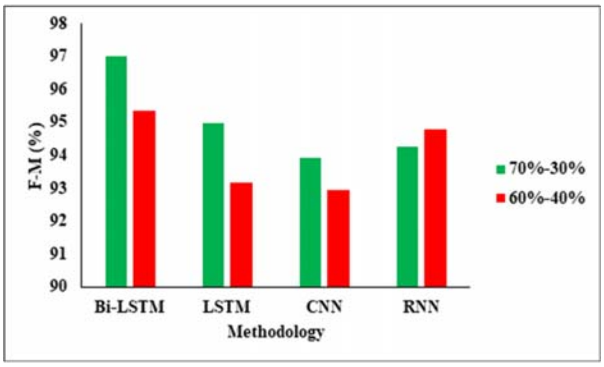

Figure 9: F-M comparison for DL techniques with optimizer

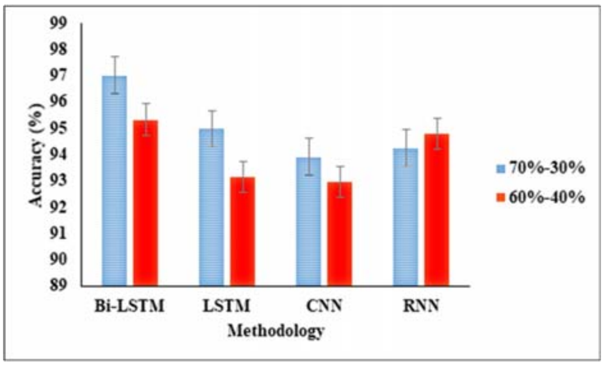

Figure 10: Accuracy comparison for DL techniques with optimizer

Here, there are two different schemes were used such as 70% Train-30% Test and 60% Train-40%Test. In the analysis of 70% Train- 30% Test scheme, Bi-LSTM reached the accuracy of 97.02%, 97% of F-M, 96.73% of S and 97.28% of P. But, the existing techniques achieved nearly 93% to 94% of accuracy, F-M, S and P. From this analysis, it is clearly mentioned that the proposed model achieved better accuracy. But, in the analysis of 60% Train-40%Test, Bi-LSTM reached accuracy of only 95.34%, where the existing models achieved nearly 91% to 93% of accuracy respectively. In this comparison analyses, the 70% Train- 30% Test data provide the better results in all classifier, which shows the importance of data. By considering the number of instances, the performance of all techniques including proposed model is tested with three metrics, which is provided in Table 3.

Table 3. Number of instance base proposed model evaluation

Sensitivity (%)				
Number of Instance	**RNN**	**CNN**	**LSTM**	**Proposed Bi-LSTM**
100	87.90	92.60	93.30	94.80
200	84.60	88.40	92.30	95.20
300	86.40	93.20	93.60	96.30
400	88.60	92.40	96.90	97.60
500	89.10	93.60	96.00	98.00
Specificity (%)				
100	83.40	84.20	92.60	94.70
200	83.60	86.10	91.20	96.80
300	86.90	87.30	92.40	95.00
400	82.10	88.30	88.60	91.20
500	86.40	89.30	90.40	93.80
Accuracy (%)				
100	76.80	89.40	91.60	95.10
200	78.60	91.30	92.40	95.90
300	77.80	87.60	90.40	95.30
400	80.10	86.40	93.20	97.10
500	82.40	86.30	92.80	97.40
F-Measure (%)				
100	88.16	89.04	89.10	93.17
200	88.97	89.67	89.76	94.61
300	89.43	90.04	92.17	94.83
400	90.28	91.62	93.15	95.22
500	91.39	92.45	95.48	96.07

In the above Table.3, the analysis evaluated the classifier model by different instance as 100-500. Also, the result performance is evaluated by using the RNN, CNN, LSTM and Proposed Bi-LSTM with optimizer. In this comparison analysis, the proposed Bi-LSTM reaches the better efficiency of 98.00% of sensitivity, 93.80% of specificity and 97.40% of accuracy at 500 instances respectively. The existing technique LSTM model achieved 96.00% of sensitivity, 90.40% of specificity and 92.80% of accuracy, CNN achieved 93.60% of sensitivity, 89.30% of specificity and 86.30% of accuracy and RNN achieved 89.10% of sensitivity, 86.40% of specificity and 82.40% of accuracy. In the analysis of F-Measure, the proposed model achieved nearly 93% to 96% for 100 to 500 instances, where the existing technique RNN achieved nearly 88% to 91%, CNN achieved nearly 89% to 92% and LSTM model achieved nearly 89% to 93% for the 100 to 500 instances. Figure 11 to 14 provides the graphical representation of various parameters with number of instances.

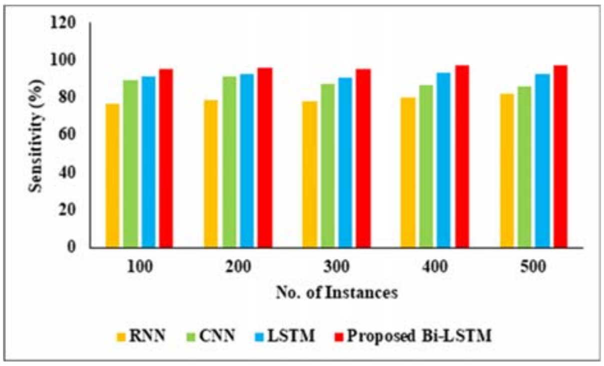

Figure 11: Sensitivity comparison with different number of instances

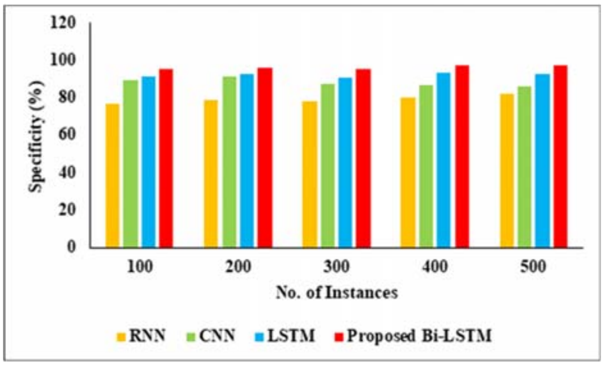

Figure 12: Specificity comparison with different number of instances

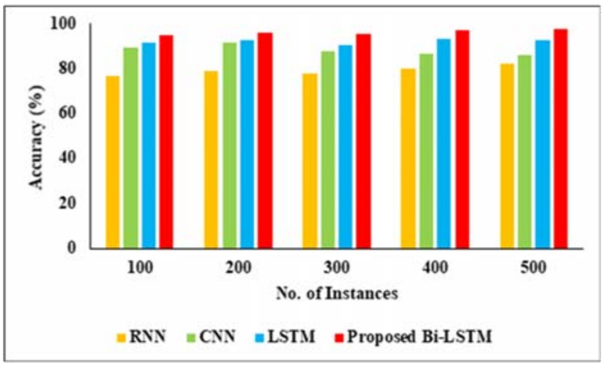

Figure 13: Accuracy comparison with different number of instances

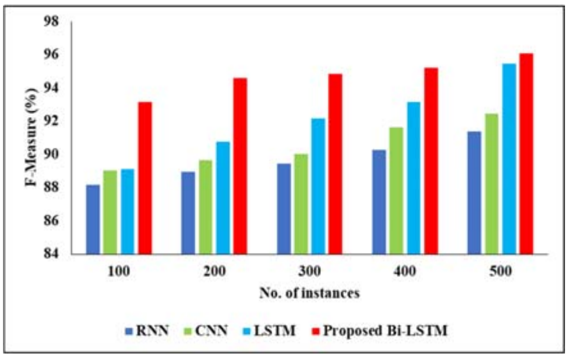

Figure 14: F-measure comparison with various number of instances

Table 4: Execution times (s) of the proposed model for various ratio of whole dataset

Architecture	70%-30%	60%-40%
RNN	119.80	112.26
LSTM	115.79	107.65
CNN	101.24	91.44
Proposed Bi-LSTM	84.27	84.04

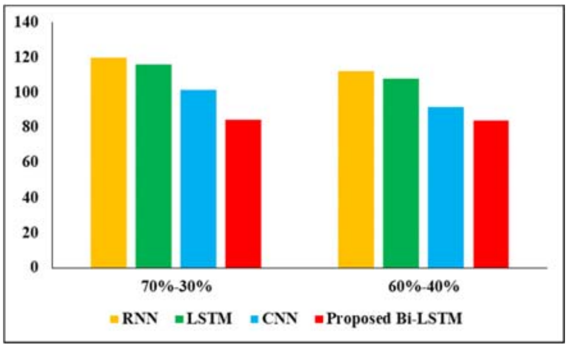

Figure 15: Graphical Comparison for Execution time

In the above Table 4 and Figure 15, the analysis of the different architecture called proposed Bi-LSTM, LSTM, CNN and RNN for execution time is provided. Here, the results are evaluated by using different data partitions as 70%-30% and60%-40%. The proposed Bi-LSTM achieved the 84.27s at 70%-30% and 84.04s at 60%-40%. But the existing techniques achieved 119.80s for RNN, 101.24s for CNN and 115.79s for LSTM, when the ratio of training data is 70% and testing data is 30%. The proposed model achieved the best results than other compared models, which is proven by the above table.

5. Conclusion

In this research work, sports interest among students is predicted by using proposed Bi-LSTM model. The data from BASS is considered as input and data augmentation is carried out during pre-processing. After that, interest prediction process is done by using bi-directional LSTM model. In this model, the hyper-parameter is optimized by using Adagrad optimizer to improve the classification accuracy of the proposed model. In the analysis process, two types of data ratio are considered that includes 70%-30% and 60%-40%. The proposed model achieved 97.02% of accuracy and 96.73% of sensitivity for 70%-30%, where the existing techniques achieved 95% of accuracy and 94% of sensitivity for the same data ratio. When the number of instances is 200, the proposed model achieved 95.90% of accuracy, where the existing LSTM model achieved 92.40% of accuracy. But when the instance is 500, the proposed model achieved 97.40% of accuracy and existing LSTM model achieved only 92% of accuracy. However, the results need to be improved by introducing biological meta heuristic algorithms for parameter tuning as a future work.

References

[1] Ma H, Pang X. Research and analysis of sport medical data processing algorithms based and Internet of Things. IEEE Access. 2019 Aug 22; Vol. 7, pp. 118839-49.

[2] Sun F. Construction Principles of Physical Fitness Training Objective System Based on Machine Learning and Data Mining. Security and Communication Networks. 2022 Apr 13; 2022. Volume 2022, Article ID 4774372.

[3] Kathiresan, N., and Samuel Manoharan, J. (2015), "A Comparative analysis of fusion- based techniques based on multi resolution transforms", National Academy Science Letters – Springer, Vol. 38, Issue.1, pp. 61 – 65.

[4] Jiang, H., Qiu, T. and DeepaThilak, K., 2021. Application of deep learning method in automatic collection and processing of video surveillance data for basketball sports prediction. Arabian Journal for Science and Engineering, pp.1-11. https://doi.org/10.1007/s13369-021-05884-1

[5] Samuel Manoharan, J. (2020), "Population based metaheuristics algorithm for performance improvement of feed forward Neural Network", Journal of Soft Computing Paradigm, Vol.2, no. 1: 36 – 46.

[6] Sivaraman, E., &Manickachezian, R. (2018). Unevenness measurement using the support vector machine and dynamic multiservice load balancing with modified genetic algorithm in cloud-based multimedia system. International Journal of Computer Aided Engineering and Technology, Vol. 10, no. 6, pp.732-747, https://doi.org/10.1504/ijcaet.2018.095210.

[7] Haiyun, Z. and Yizhe, X., 2020. Sports performance prediction model based on integrated learning algorithm and cloud computing Hadoop platform. microprocessors and microsystems, Vol. 79, p.103322.

[8] Liu X, Khadidos AO, Keir MY. Research on predictive control of students' performance in PE classes based on the mathematical model of multiple

linear regression equation. Applied Mathematics and Nonlinear Sciences. 2021 Nov 22. DOI: https://doi.org/10.2478/amns.2021.2.00058

[9] Li B, Xu X. Application of artificial intelligence in basketball sport. Journal of Education, Health and Sport. 2021 Jul 8; Vol.11(7):54-67.

[10] Horvat T, Job J. The use of machine learning in sport outcome prediction: A review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2020 Sep; Vol. 10, no. 5, e1380.

[11] Hou S. Research on the Application of Data Mining Technology in the Analysis of College Students' Sports Psychology. Mobile Information Systems. 2021 Nov 23;2021. Volume 2021, Article ID 6529174.

[12] Xu W, Xiong W, Shao Z, Li Y. Analysis of effectiveness and performance prediction of sports flipped classroom teaching based on neural networks. Scientific Programming. 2021 Jul 31;2021. Volume 2021, Article ID 5284457.

[13] Sokkhey P, Okazaki T. Developing web-based support systems for predicting poor-performing students using educational data mining techniques. International Journal of Advanced Computer Science and Applications. 2020; Vol. 11, no. 7.

[14] Yang J, Chen M. Construction of Sports and Health Data Resources and Transformation of Teachers' Orientation Based on Web Database. Journal of Healthcare Engineering. 2022 Feb 7;2022.Volume 2022 Article ID 4372406.

[15] Adebisi J, Alabi D. Design and Implementation of a Predictive Model for Nigeria Local Football League. International Journal of Computer Science and Security (IJCSS). 2021 Aug 1; Vol. 15, no. 4:106-23.

[16] Bunker R, Susnjak T. The application of machine learning techniques for predicting results in team sport: a review. arXiv preprint arXiv:1912.11762. 2019 Dec 26.

[17] Roy KS, Roopkanth K, Teja VU, Bhavana V, Priyanka J. Student career prediction using advanced machine learning techniques. International journal of engineering and technology. 2018; vol. 7(2.20).

[18] Haiyun Z, Yizhe X. Sports performance prediction model based on integrated learning algorithm and cloud computing Hadoop platform. microprocessors and microsystems. 2020 Nov 1; Vol. 79:103322.

[19] Sorkkila M, Aunola K, Salmela-Aro K, Tolvanen A, Ryba TV. The co-developmental dynamic of sport and school burnout among student-athletes: The role of achievement goals. Scandinavian journal of medicine & science in sports. 2018 Jun; vol. 28(6):1731-42.

[20] Zhang X, Lou J. Performance Prediction and Action Performance Analysis of Sports Competitive Events Based on Deep Learning. Mathematical Problems in Engineering. 2022;2022. Volume 2022, Article ID 9436736

[21] Ting J, Chien KP, Dhir A, Chen S. Sports interest mediating exercise and compulsive internet use among undergraduates. Health promotion international. 2019 Oct 1; Vol. 34, no. 5, pp. 953-60.

[22] A Basheer Ahamed, M. Mohamed Surputheen. Real Time Sports Distribution Analysis Model for Improved Student Interest Prediction in Sports Mining. international Journal of Advanced Science and Technology Vol. 29, No. 5s, (2020), pp. 513-522.

[23] Ma B, Nie S, Ji M, Song J. Research and analysis of sports training real-time monitoring system based on mobile artificial intelligence terminal. Wireless Communications and Mobile Computing. 2020 Sep 28;2020. Volume 2020 Article ID 8879616

[24] Zhang R. College sports decision-making algorithm based on machine few-shot learning and health information mining technology. Computational Intelligence and Neuroscience. 2022 Mar 31;2022. Volume 2022 Article ID 7688985

Mr. A. Basheer Ahamed, a Research Scholar at Jamal Mohamed College (Autonomous) cleared his State Level Test (TNSET-2016) conducted by Mother Teresa University, Kodaikanal.He has completed his Master of Philosophy (C.S) in 2007 and Master of Computer Science in 2000. He is currently working as an Assistant Professor of Computer Science at Jamal Mohamed College for 14 Years. His areas of research include Data mining, Machine learning and Deep Learning.

Communication maid id: abammb2@gmail.com

Dr. M. Mohamed Surputheen, Working as an Associate Professor in Jamal Mohamed College with more than 30yrs experience. Education Qualification is M.Sc.. M.Phil and PhD. He is Guiding M.Phil Scholars and around eight PhD Scholars. Published and presented more than twenty five research papers. Currently acting as a Coordinator of ComputerScience and IT Department at Jamal Mohamed College (Autonomous), Trichy. His areas of research include Wireless sensor networks, Data mining, Machine Learning, Deep Learning and Image Processing.

Communication mail id: msurfudeen@yahoo.com