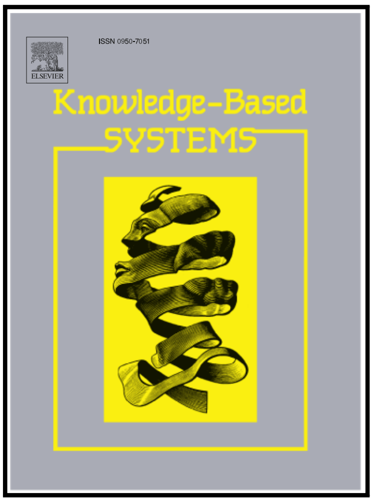
Journal Pre-proof



Predicting Student Satisfaction in Metaverse Platform Using Deep Learning

Shaik Shameen Taz , B. Narendra Kumar Rao

PII: S0950-7051(26)00694-5
DOI: https://doi.org/10.1016/j.knosys.2026.115968
Reference: KNOSYS 115968

To appear in: *Knowledge-Based Systems*

Received date: 19 January 2026
Revised date: 6 April 2026
Accepted date: 7 April 2026

Please cite this article as: Shaik Shameen Taz , B. Narendra Kumar Rao , Predicting Student Satisfaction in Metaverse Platform Using Deep Learning, *Knowledge-Based Systems* (2026), doi: https://doi.org/10.1016/j.knosys.2026.115968

Predicting Student Satisfaction in Metaverse Platform Using Deep Learning

[1*]Shaik Shameen Taz and [2]Dr. B. Narendra Kumar Rao

[1*]Research scholar, School of Computing

Mohan Babu University

Tirupati, Andhra Pradesh,517102, India.

*shaikshameen92.t@gmail.com

[2]Professor, School of Computing

Tirupati, Andhra Pradesh,517102, India.

Abstract: Predicting student satisfaction on Metaverse platforms can improve educational effectiveness by analyzing factors like engagement, interaction, and content delivery. The benefits include personalized, immersive learning experiences and real-time feedback. However, the existing schemes suffer from overfitting issues due to highly imbalanced data and fail to generalize to diverse content. To address these limitations, a robust model, Deep High Attention Long Short-Term Memory Forward Harmonic Network (DHA-LSTM-FH Net), is proposed to predict student satisfaction in Virtual Reality (VR) teaching within the Metaverse. The model utilizes data from Metaverse platforms, including virtual spaces, Augmented Reality (AR)/VR devices, learning materials, and student information. Initially, interaction logs from VR sessions and student profiles are collected as input. The data undergoes softmax normalization to ensure consistency. Feature selection is conducted using Recursive Feature Elimination (RFE) and Elastic Net to select the key features. Local Density-based Synthetic Minority Over-Sampling Technique (LD-SMOTE) is then applied to address data imbalance. Student satisfaction prediction is done by DHA-LSTM-FH Net, which is developed by combining Deep High Attention Neural Network (DHA-Net) and Long Short-

Term Memory (LSTM) using Harmonic Analysis. Experimental results show that the model achieves a precision of 93.765%, a recall of 95.755%, an F1 Score of 94.750%, and a Cohen's Kappa of 0.838, outperforming baseline methods. However, the model is trained on a specific VR/Metaverse platform, so its performance may drop when applied to different Metaverse setups or content types.

Keywords: Metaverse, Deep Learning, Virtual reality, Student satisfaction prediction, Long Short-Term Memory

1. Introduction

Advances in computer science have greatly changed daily life by revolutionizing communication, social interaction, and economic activities. From the perspective of end users, three major waves of technological advancement have occurred: the rise of personal computers, the emergence of the Internet era, and the proliferation of mobile devices. Currently, the experience of the fourth major wave is characterized by immersive and spatial technologies, including Augmented Reality (AR) and VR [1]. These emerging tools are expected to shape the next era of ubiquitous computing, with the potential to transform key sectors such as education, business, entertainment, and remote work. This emerging paradigm is referred to as the Metaverse. The term "Metaverse" is a compound of "meta," meaning virtual or transcendent [2], and "verse," which denotes the universe or world. The term 'Metaverse,' which originally referred to an online virtual world, has grown in the post-pandemic era to serve as a connector between the online and offline spheres [3]. The Metaverse now represents a continuous, immersive, multi-user digital ecosystem that blends the real world with virtual experiences [4]. Within the realm of online education, the Metaverse offers promising solutions to overcome the inherent limitations of traditional 2D web-based educational platforms. Despite technological progress, the core structure of education has remained relatively unchanged, still heavily reliant on textbooks, classrooms, and content delivery [5]. However, within a Metaverse learning space, students can engage more deeply with their peers through avatars, contribute creative work in the form of 3D models, and collaborate via virtual social networks. This fosters a more interactive and communal learning experience, supporting the evolution of digital collaborative education [6] [7].

VR is a rapidly developing technology that creates interactive, computer-generated simulations, offering users the sensation of being present in real-time environments. By integrating advanced computer graphics, motion tracking, and immersive display systems, VR in education enables learners to interact with real-world environments without leaving the classroom [8]. Although VR is widely recognized for its impact on entertainment and gaming, its applications extend far beyond these industries. It has become an essential tool in fields like real estate, engineering, healthcare, and, increasingly, education [1] [9]. In education, VR enables students to interact with both real and conceptual environments, offering immersive experiences that extend beyond the limits of traditional teaching. The versatility of VR has made it a valuable educational resource across a broad range of subjects. It has already been integrated into teaching disciplines like medicine, biology, chemistry, physics, astronomy, mathematics, psychology, architecture, history, and engineering [10] [11]. The versatility of VR in education is evident, with applications in science, history, architecture, and archaeology, among others. A major benefit of virtual reality in education is its ability to deliver information through immersive, experiential learning rather than through traditional descriptive methods. Students can engage directly with content that would otherwise be difficult or even impossible to demonstrate through textbooks or conventional teaching tools. From exploring ancient ruins in archaeology to visualizing scientific phenomena that are invisible to the naked eye, VR opens new pathways for interactive and immersive learning [12].

Predicting student satisfaction has emerged as a significant application area of Deep Learning (DL), offering promising opportunities to enhance academic performance and support personalized learning experiences [13] [14]. The rise of DL, alongside the growth of big data science, has contributed greatly to the success of many leading companies. Its use extends far beyond the academic sphere, finding impactful applications across diverse domains such as healthcare, cybersecurity, banking, robotics, and retail. In various tasks, such as image classification, action detection, NLP, signal processing, and linguistic communication, DL systems have surpassed both traditional methods and human benchmarks [15] [16]. Their strength lies in the ability to perform advanced classification and prediction tasks across complex datasets [13] [17]. DL is a subfield of Machine Learning (ML) that uses multi-layer Artificial Neural Networks (ANNs), enabling models to automatically learn hierarchical feature representations from input data [18] [19] [20] [21] [22] [23]. This capability is particularly valuable when working with large volumes of unlabeled or high-dimensional data, enabling deeper insights and more accurate predictions [24] [25]. Technologies like CLIP highlight the growing shift from traditional ML systems to more dynamic DL frameworks, pushing the boundaries of what machines can learn and understand [26]. These advances have far-reaching implications, especially in education, where DL-powered Artificial Intelligence (AI) systems are being integrated into Edu-metaverse platforms to create more interactive and immersive learning experiences. However, this evolution also introduces new challenges that must be thoughtfully addressed [27].

Predicting student satisfaction in VR-based teaching within the metaverse poses several challenges. Key issues include accurate data collection, varying levels of student engagement, technical limitations such as latency, the need to maintain immersive learning experiences, and the requirement to support diverse learning preferences. In addition, the dynamic and continuously evolving nature of VR platforms makes the development of consistent and reliable prediction models more complex. However, they often encounter overfitting issues, and their computational complexity is high [28]. These challenges highlight the necessity for an efficient and robust prediction framework capable of handling complex interaction data generated in metaverse-based educational environments. To address these issues, an innovative technique termed DHA-LSTM-FH Net is proposed for predicting student satisfaction with VR-based teaching in the Metaverse. The developed methodology begins with the collection of interaction logs from VR sessions, along with relevant student information, from the Metaverse educational platform that integrates virtual environments, AR/VR devices, learning materials, and user data. Further, a sigmoid function is applied to ensure data consistency and comparability. Feature selection is then performed using techniques such as Elastic Net and RFE to identify the most significant attributes. To expand and diversify the dataset, the process of data augmentation is implemented using LD-SMOTE. Student satisfaction in VR-based teaching within the metaverse is predicted using a DL model known as DHA-LSTM-FH Net. This model is introduced by combining DHA-Net and LSTM with harmonic analysis techniques.

The primary contribution of this research is provided as follows,

- **Deep High Attention Long Short-Term Memory Forward Harmonic Network to Predict Student Satisfaction in VR Teaching within the Metaverse:** To predict student satisfaction in VR-based teaching within the metaverse, the DHA-LSTM-FH Net model is used, which merges DHA-Net, LSTM through the application of Harmonic Analysis. Moreover, Feature selection is conducted using techniques such as ElasticNet and RFE to identify the most significant attributes.

1.1 System model of the metaverse education platform

A Metaverse education platform integrates 3D virtual environments, AR/VR hardware, digital learning materials, and student information to provide immersive and interactive learning experiences. Students engage in virtual spaces, such as classrooms and labs, using AR/VR devices that enable hands-on interaction with educational content and support collaborative participation. All student activity is tracked through access logs, capturing engagement patterns and interaction data, which enable educators to monitor performance and adapt instruction. The platform also manages student profiles, organizes assignments, and integrates learning materials like lesson plans, simulations, and interactive exercises, supporting diverse educational objectives. By combining these components, the system enables personalized learning, data-driven insights, and adaptive feedback. Figure 1 illustrates the overall system model, highlighting the flow from data collection to predictive analysis and satisfaction output.

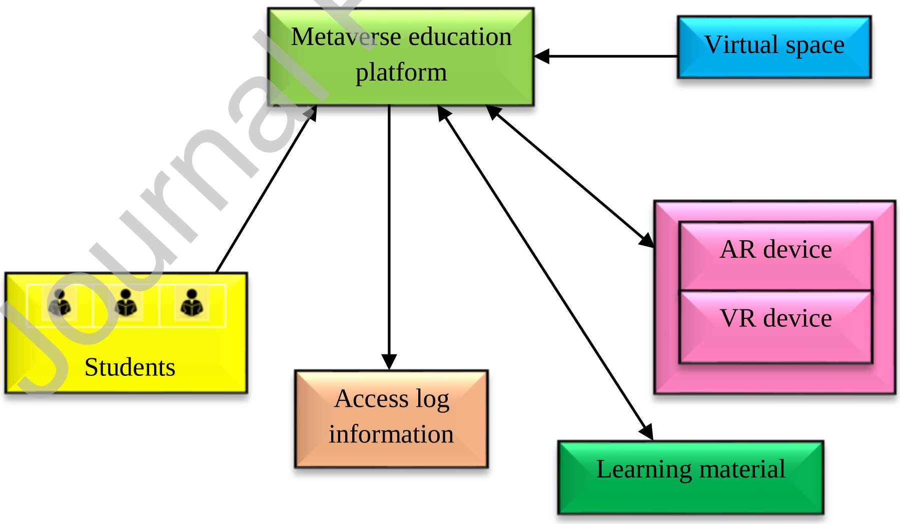


Figure 1. The system model of the Metaverse education platform

The structure of the rest of this paper is outlined as shown below: In Section 2, the current approaches to predicting student satisfaction in VR teaching environments within the

Metaverse are explored. Section 3 provides a thorough description of the DHA-LSTM-FH Net model. Section 4 presents the experimental setup and analyzes the results. Finally, Section 5 concludes the paper by summarizing key points and proposing directions for future exploration.

2. Literature survey

Rahman, H., *et al.* [2] developed an immersive VR-based platform called the Virtual Chemistry Classroom for Chemical Bonding (VC3B), designed to enhance the learning of chemical bonding and formulas through a game-based learning approach. This method proved effective in increasing student engagement and demonstrated practical applicability in real-world educational scenarios. However, this model could not address the more complex and sophisticated chemistry principles that are necessary at higher levels of education. Arpaci, I. and Bahari, M., [6] introduced a hybrid method that combined Covariance-Based Structural Equation Modelling (CB-SEM) with a Deep ANN to predict the educational sustainability of the Metaverse. This approach improved prediction accuracy by effectively capturing complex patterns in the data. Nonetheless, this approach did not contribute to the advancement of existing theories or knowledge within the Information Systems literature. Ali, R.A., *et al.* [29] developed a hybrid approach combining Partial Least Squares Structural Equation Modelling (PLS-SEM) with an ANN model to utilize the metaverse in higher educational institutions. This method proved to be computationally efficient and reliable. Nevertheless, this model failed to detect the precise areas of student interest and the critical determinants influencing their adoption of novel technologies. The Structural Equation Modelling (SEM) with ML algorithm was designed by Almarzouqi, A., *et al.* [30] to predict users' intentions to adopt metaverse systems in medical education. This approach effectively avoided overfitting and achieved high precision. However, the applicability of the technique was limited, as the model focused solely on educational contexts, thereby reducing its generalizability to other domains.

Alshammari, S.H. and Alshammari, M.H., [31] introduced a Structural Equation Modelling (SEM) approach to examine their impact on students' intention to use the metaverse in higher education. This method was noted for its robustness and ease of accessing learning-related information. However, the sample used in the model lacked geographic and demographic diversity, which constrained the external validity and the generalizability of the results. Akour, I.A., *et al.* [32] developed a hybrid SEM-ANN approach to examine metaverse adoption in higher education institutions in the Gulf region. This model effectively minimized overfitting and reduced the error rate in estimating the ability to predict the behavioral intention to utilize the metaverse system. Nevertheless, its application was restricted to educational contexts, limiting its adaptability to other domains outside the influence of metaverse-based teaching and learning systems. PLS-SEM and a fuzzy-set Qualitative Comparative Analysis (fsQCA) model were designed to investigate factors influencing Egyptian university students continued use of the metaverse as a learning platform. The approach was computationally efficient, enhanced learning outcomes, and improved information accessibility. Although the technique demonstrated notable strengths, the exclusion of participants from both private and public universities compromised the ability to perform a comparative analysis, thereby limiting the generalizability of the findings. The Technology Acceptance Model (TAM) was proposed by Al-Adwan, A.S., *et al.* [33] to explore university students' intentions to use metaverse-based learning platforms. This method was flexible and involved lower computational complexity. However, this method was not well-suited for real-time scenarios, reducing its effectiveness in dynamic or live educational settings.

Manikandan Suriyanarayanan., *et. al* [34] introduced the use of an ANN–based predictive modelling approach to assess student satisfaction in online learning environments. Unlike regression-based models, ANN can capture complex, nonlinear interactions among variables such as behavioral intention, attitude, and satisfaction. Yet, the ANN models require large, clean, and well-preprocessed datasets. Sulis Sandiwarno., *et. al* [35] introduced a novel DL model called E-learning Satisfaction Assessment using Textual Neural Network (E-SATNet) by integrating sentiment and semantic similarity analysis using Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) for student satisfaction in online learning environments. This dual-branch design captures both emotional tone and thematic alignment, giving a more complete picture of student satisfaction. However, the dual-branch architecture requires significant computational resources; also, training and inference are more expensive compared to simpler models. Nisar Ahmed Dahri., *et.al* [36] developed a multimodal ML technique, using XGBoost, RF, and SVM, with XGBoost emerging as the best-performing model, and an explainable AI (XAI) framework. By integrating XAI, this approach ensured accountability and built trust among stakeholders. While SHAP and LIME improve transparency, they provide local explanations that may not fully capture global model behavior.

2.1 Challenges

The challenges associated with techniques for predicting student satisfaction in the metaverse are provided below,

- The Hybrid CB-SEM and Deep ANN Model [6] effectively identified and modelled complex relationships between constructs, offering actionable insights for enhancing virtual learning platforms in the Metaverse. However, this model did not succeed in fostering greater student engagement, a crucial element for boosting satisfaction, as students tend to have more fulfilling experiences when actively involved in their learning.
- The PLS-SEM with ANN approach [29] provided a strong predictive framework for assessing ongoing engagement in Metaverse environments. Nevertheless, the model struggled to mitigate the impact of outliers or extreme data points, thereby limiting its

effectiveness in accurately forecasting outcomes such as user retention and overall satisfaction.

- The Hybrid SEM-ML approach [30] underscored the potential of Metaverse technologies to enrich learning, drive engagement, and facilitate collaboration. Despite these strengths, the model could not automatically eliminate irrelevant variables, particularly in situations involving multicollinearity, which can distort results.
- The TAM model [33], incorporating personal innovativeness and service quality, showed increased relevance for higher education adoption of Metaverse technologies. However, this model failed to capture complex data patterns and was unable to generate realistic synthetic data samples for more refined predictions.
- Metaverse in education is relatively new; there is a scarcity of mature, standardized frameworks to accurately assess student satisfaction and learning outcomes. Traditional models such as TAM and Self-Determination Theory (SDT) may not fully reflect the nuanced dynamics of immersive, VR-based learning experiences in the Metaverse.

3. Proposed Deep High Attention Long Short-Term Memory Forward Harmonic Network to predict student satisfaction

The main goal of this work is to design a robust and accurate predictive approach, known as DHA-LSTM-FH Net, for predicting student satisfaction with VR-based teaching in the Metaverse education platform. The proposed methodology utilizes data from the Metaverse educational platform that incorporates virtual spaces, AR/VR devices, learning materials, and student-related information. The initial step involves gathering interaction logs from VR sessions, along with corresponding student data, which serve as the primary input. To ensure consistency and comparability, the collected data undergo data normalization utilizing a sigmoid function [37]. Subsequently, feature selection is performed using techniques, such as

Elastic Net [38] and RFE [28], to identify the most influential attributes affecting student satisfaction. Furthermore, LD-SMOTE [39] is applied for data augmentation to address data imbalance. Thereafter, student satisfaction is predicted using the developed DL model, named DHA-LSTM-FH Net. This model integrates the DHA-Net [40] and LSTM [41] using Harmonic Analysis [42]. In Figure 2, the schematic representation of the proposed DHA-LSTM-FH Net approach for predicting students' satisfaction is illustrated.

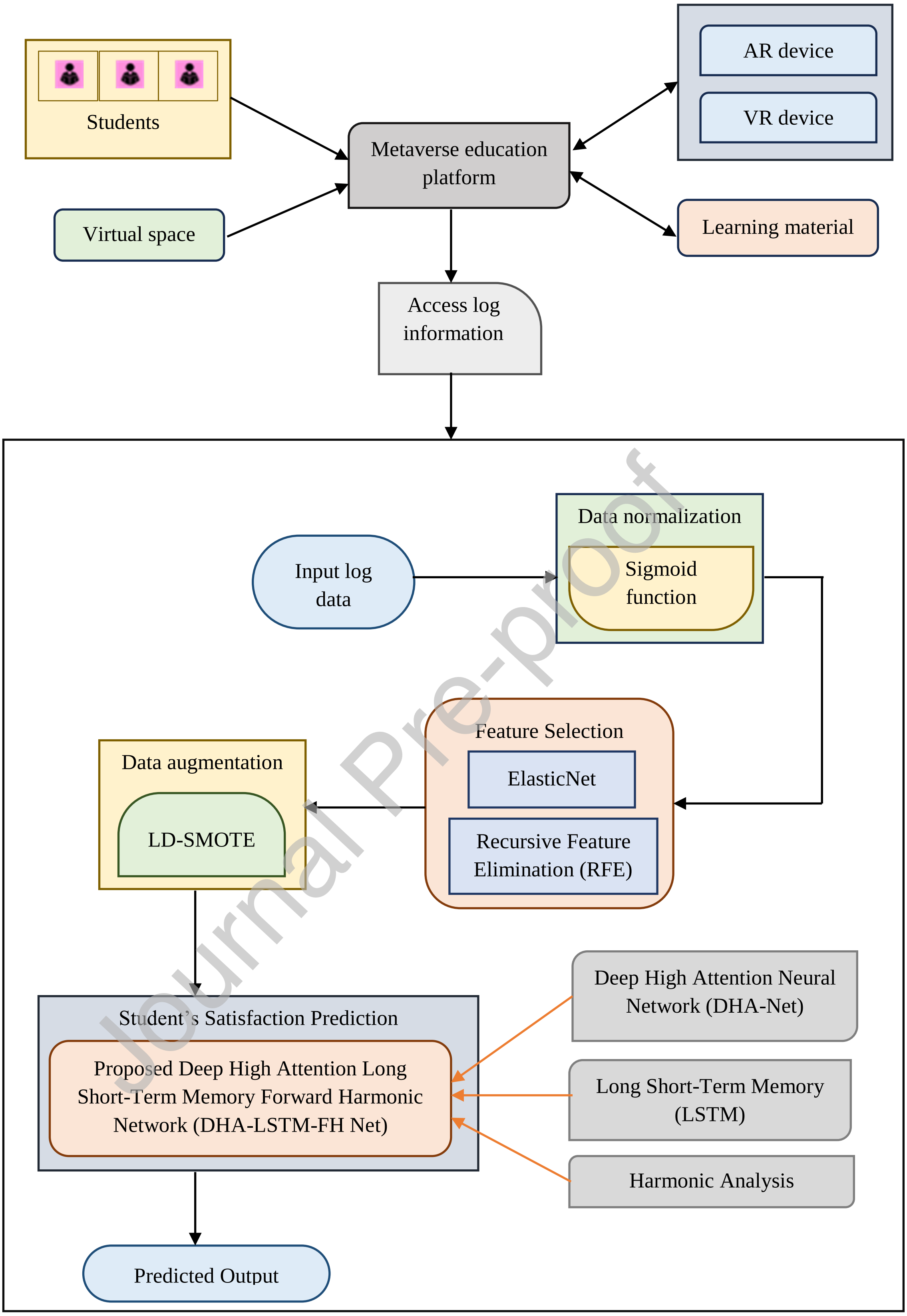


Figure 2. Block diagram of the developed DHA-LSTM-FH Net for predicting students' satisfaction

3.1 Acquisition of data from the Metaverse Education platform

The students' satisfaction prediction is accomplished based on the datasets, such as the Impact of Virtual Reality on Education dataset [43] and VR_College_Teaching_Dataset [44]. The dataset exemplified as C is formulated by,

$$C = \{C_1, C_2, C_3, \ldots, C_u, \ldots C_v\} \tag{1}$$

wherein, in the dataset, the u^{th} data is delineated as C_u, as well as the entire number of data records is specified as v.

3.2 Data Normalization utilizing Sigmoid Function

Data normalization is a technique employed to adjust data to a standard scale without distorting the inherent differences in the range of values. In the context of predicting student satisfaction in the metaverse, normalization plays a crucial role in minimizing biases arising from varying scales of input data, such as student demographic profiles, VR usage metrics, and feedback from students and educators. By normalizing the data, the model's accuracy is improved, training is accelerated, and consistent comparisons are ensured across different features. As a result, normalization enhances the model's ability to make better predictions and generate more insightful outcomes. In this work, the sigmoid function [37] is employed for normalization. The sigmoid function is a nonlinear transformation that maps input values into a bounded range between 0 and 1, making it suitable for scaling heterogeneous features into a common range. It is applied to the input data C_u, and the normalized data $\sigma(z)$ is produced as the output.

Sigmoid function

The sigmoid function, also known as the logistic function, is a nonlinear activation function widely used in ML and DL for transforming input values into a bounded range between 0 and 1. It is particularly useful in normalization and binary classification tasks due to its smooth and differentiable nature. The choice of the sigmoid function is motivated by its ability to smoothly scale input data while preserving relative differences between feature values. Unlike softmax, which operates on multi-dimensional vectors to produce a probability distribution, the sigmoid function is applied element-wise, making it more suitable for feature-wise normalization in this study. Additionally, sigmoid normalization helps in stabilizing gradient updates, reducing the impact of extreme values, and improving the learning capability of deep neural networks. Thus, sigmoid-based normalization ensures that all input features are transformed into a uniform range, facilitating efficient training of the proposed DHA-LSTM-FH Net model. The sigmoid function is defined as,

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{2}$$

here, $\sigma(z)$ epitomizes the normalized feature value, and the input to be scaled is epitomized as z.

3.3 Feature selection with RFE and Elastic Net

Feature selection refers to the process of identifying and selecting the most relevant variables from a dataset for use in predictive modelling. This process enhances model performance by increasing accuracy, reducing overfitting, and improving interpretability. Feature selection enables more streamlined and insightful models. In this scenario, the normalized data $\sigma(z)$ is fed as input to the feature selection module, where methods, like Elastic Net [38] and RFE [45], are used to select the set of significant features H as an output. Traditional feature selection methods or simple models often fail to handle correlated features, may select irrelevant variables, and can lead to overfitting, especially with high-dimensional datasets. In contrast, RFE iteratively removes the least important features, ensuring the model focuses on the most influential variables, while Elastic Net combines L1 and L2 regularization to handle correlated features and promote sparsity. Together, this combination selects a robust and meaningful feature set, improving accuracy, generalization, and interpretability, which makes it particularly suitable for predicting student satisfaction from complex Metaverse interaction data.

3.3.1 Elastic Net

ElasticNet [38] is a regularized linear regression model that combines Z_1 (Lasso) and Z_2 (Ridge) penalties, making it effective for feature selection. It manages multicollinearity well and tends to select groups of correlated features. The normalized input data $\sigma(z)$ is provided as input to the ElasticNet, and the selected feature set H_1 is produced as output. This approach improves prediction performance, enhances robustness against overfitting, and is particularly suited for high-dimensional, sparse datasets.

Elastic Net [38] is an extension of multiple linear regression techniques, specifically developed for handling problems with high-dimensional feature selection. It combines two penalty terms, such as the $L_1 - norm$ and $L_2 - norm$ to automatically select relevant variables and apply continuous shrinkage. This dual-penalty approach enhances prediction accuracy by retaining important predictors (features) while discarding irrelevant ones. The method is often likened to a flexible fishing net, as it captures the "big fish" (i.e., key features) while letting insignificant variables slip through.

Consider a dataset with q predictors specified as $d_1,\ldots,d_q$. Using linear regression, the response variable t is modelled as: $\hat{t} = \delta_0 + \delta_1 d_1 + \ldots + \delta_q d_q$, where the coefficients $\left(\delta = \left[\delta_0, \ldots \delta_q\right]\right)$ are estimated by minimizing the sum of the squared error residuals. The Sum of Squared Errors (SSE) in a regression model is expressed as,

$$SUM = \left\| T - D\delta \right\|^2 \qquad (3)$$

In the above equation, SSE is typified as SUM, the actual value is epitomized as T, the regression coefficient is indicated as δ, and the predicted value is exemplified as D.

When the count of features surpasses the observations count, coefficient estimation requires the minimization of an alternative L function instead of the SSE, with the overall loss being expressed as follows,

$$L = SUM + \chi\rho\|\delta\|_1 + \chi(1-\rho)\|\delta\|^2 \quad (4)$$

wherein, χ and ρ indicates the weighting terms.

In this context $\|\delta\|_1$ and $\|\delta\|^2$ are computed as,

$$\|\delta\|_1 = \sum_{q=1}^{Q} |\delta_q| \quad (5)$$

$$\|\delta\|^2 = \sum_{q=1}^{Q} \delta_q^2 \quad (6)$$

here, the overall number of features is represented as Q. The Elastic Net regression model includes a regularization process where model complexity is controlled by two weighting parameters χ and ρ. These parameters must be tuned during the learning process, as they significantly influence the model's outcome. Elastic Net includes two special cases, when the weighting parameter is set to 1 $(\rho = 1)$, the model becomes equivalent to the Least Absolute Shrinkage and Selection Operator (LASSO) regression, which encourages sparsity by driving many coefficient values to zero. When the weighting parameter is 0 $(\rho = 0)$, the model reduces to ridge regression, which shrinks the coefficients of correlated predictors together rather than eliminating them, helping to include more features. Elastic Net combines the strengths of both ridge regression and LASSO. The selected feature obtained from the Elastic Net approach is epitomized as H_1.

3.3.2 Recursive feature elimination

Recursive Feature Elimination (RFE) [45] is a feature selection technique that systematically removes the least important features based on model performance. The process involves fitting a model, ranking features according to their importance, and iteratively eliminating the least significant features until the optimal subset is obtained. This method offers several advantages, including improved model accuracy, reduced risk of overfitting, and enhanced computational efficiency by discarding irrelevant or redundant features. The normalized data $\sigma(z)$ is fed as input to the approach, which results in the selected feature set H_2 as an outcome.

RFE [45] is a feature selection technique that operates as a wrapper method through a backward elimination process. Initially, the approach is trained employing the complete sequence of features. After ranking the features according to their importance, the one with the least influence on training error is eliminated, and the model is retrained to assess performance. The cycle repeats until the ideal subset of features is identified. RFE focuses on retaining the most informative and independent features, discarding unnecessary or irrelevant ones to enhance model generalization. Consider X typifies the sequence used to store feature rankings throughout the backward feature elimination process. During every iteration, a subset of top-ranked features, denoted by X_w, is saved. The model is refitted using X_w, and its performance is evaluated. After computing the value of X_w with the highest performance, the final model is fitted using the top-performing features. In Algorithm 1, the Algorithm of RFE with cross-validation is illustrated.

Algorithm 1. Algorithm of RFE with cross-validation
Apply a 10-fold cross-validation to train the logistic regression method utilizing the complete feature set
Evaluate the performance of the model
Determine the importance or ranking of all features
For every subset X_w, where $w = 0,1,2,3,\ldots\ldots,k$, perform the following:
Select the X_w features with the highest importance
Test and train the approach using the X_w features
Recompute the performance of the approach
Re-estimate the importance scores of all features
End
Evaluate the approach's performance on X_w
Identify the ideal feature count
Apply the model using the chosen optimal features

The selected feature attained from the RFE approach is signified as H_2.

The final selected feature H, obtained by merging the Elastic Net and RFE features, and is formulated as,

$$H = \wp * H_1 + (1 - \wp) H_2 \tag{7}$$

where, the result obtained from the Elastic Net approach is indicated as H_1, the final features selected are denoted as H, the parameter coefficient is exemplified as $\wp$, and the output attained from the RFE technique is represented as H_2.

3.4 Data augmentation utilizing LD-SMOTE

Existing approaches for handling imbalanced datasets, such as random oversampling or undersampling, often suffer from significant drawbacks. Random oversampling can lead to overfitting, while undersampling may discard valuable information from the majority class, reducing model accuracy. To overcome these limitations, the proposed study employs LD-SMOTE [39] for data augmentation. LD-SMOTE generates synthetic samples for minority classes while preserving the original data distribution and improving class balance, ensuring that the model does not become biased toward the majority class. In this process, the selected feature H is subjected to data augmentation using LD-SMOTE, which generates the augmented output W. The LD-SMOTE approach preserves the data distribution, improves class balance, and reduces overfitting, resulting in more robust and accurate models.

3.4.1 Local Density Estimation-Based Synthetic Minority Oversampling Technique

LD-SMOTE [39] is a data augmentation method designed to enhance the performance of ML models on imbalanced datasets. It achieves this by generating synthetic minority class samples using local density information. The framework consists of three main steps: Initially, feature contributions are calculated using an innovative distance metric within Jaccard similarity. Next, local density is estimated around each minority (positive) sample. Finally, synthetic samples are generated by forming triangles that connect a target minority sample with two of its nearest neighbors and interpolating new points within these triangles to enhance data diversity.

Assume $U = \left\{ \mathrm{M}_c, \mathrm{N}_c \right\}_{c=1}^{a}$ represents a binary classification dataset with class imbalance, and every M_c is the c^{th} feature vector with B-dimensions, and N_c exemplifies its label with $\mathrm{N}_c = +1$ for a positive class sample and $\mathrm{N}_c = 0$ for a negative class sample. The set of features is denoted as $\left\{ l_1, l_2, \ldots, l_B \right\}$ while the sequence of negative and positive instances is epitomized as M_- and M_+.

In this step, k-means clustering is applied to each feature $l_z, z = 1, 2, \ldots, B$ separately to group the training data into two clusters, with the clustering process based on the values of the feature. generating,

$$F_1^{l_z} = \left\{ \left(\mathrm{M}_c, \mathrm{N}_c \right) \mid \left(\mathrm{M}_c, \mathrm{N}_c \right) \text{belongs to the } 1^{st} \text{ cluster as determined by its features } l_z \right\} \quad (8)$$

and

$$F_2^{l_z} = \left\{ \left(\mathrm{M}_c, \mathrm{N}_c \right) \mid \left(\mathrm{M}_c, \mathrm{N}_c \right) \text{belongs to the } 2^{nd} \text{ cluster as determined by its features } l_z \right\} \quad (9)$$

For every feature l_z, the Jaccard similarity $\left(J_1\right)$ is subsequently assessed on comparison with the positive class M_+ and the clusters R_1 and R_2, which is expressed as,

$$J_1(R_1, R_2) = \frac{\left|R_1 \cap R_2\right|}{\left|R_1 \cup R_2\right|} \quad (10)$$

Hence, the similarity for each feature l_z is evaluated as,

$$J_1\left(\mathrm{M}_+, F_b^{l_z}\right) = \frac{\left|\mathrm{M}_+ \cap F_b^{l_z}\right|}{\left|\mathrm{M}_+ \cup F_b^{l_z}\right|}, b = 1, 2 \quad (11)$$

Based on the Jaccard score derived from the above equation (11), the cluster $F_b^{l_z}$, for $b \in \{1, 2\}$ that achieves the highest value is regarded as the one most similar to the positive class M_+, and is characterized as,

$$F_{cl}^{l_z} = \arg\max_{b \in \{1,2\}} J_1\left(\mathrm{M}_+, F_b^{l_z}\right) \quad (12)$$

In the above equation, the $F_{cl}^{l_z}$ indicates the cluster closest to M_+. The Jaccard similarity among $F_{cl}^{l_z}$ and M_+ is then determined by,

$$\beta_z = J_1\left(\mathrm{M}_+, F_{cl}^{l_z}\right) \quad (13)$$

In the above equation, the Jaccard score β_z, is known as the degree of z^{th} feature in classification contribution, which measures the effectiveness of that feature in sorting the training data into two different classes. Consequently, a novel metric, referred to as distance, is defined by the following expression,

$$x_{g,h} = \frac{1}{B} \sum_{z=1}^{B} \frac{\left|\mathrm{M}_{g,z} - \mathrm{M}_{h,x}^1\right|}{\beta_z} \quad (14)$$

wherein, the term $\left|\mathrm{M}_{g,z} - \mathrm{M}_{h,x}^1\right|$ corresponds to the absolute difference of the z^{th} feature for the positive samples M_g and M_h, and the overall count of features is delineated as B.

Local Density Estimation

The feature vector M_c is considered for the c^{th} sample within the sequence M_+. Initially, distance $x\left(\mathrm{M}_c, \mathrm{M}_z\right)$ is calculated between M_c and all other samples M_z using the above equation (14). Subsequently, the k nearest neighbors of M_c, are identified based on the distance. Finally, the average distance between M_c, K_z signifies the sequence of k nearest neighbors and its k nearest neighbors are evaluated as given below,

$$\eta_c = \frac{1}{k}\sum_{\mathrm{M}_z \in K_z} x\left(\mathrm{M}_c, \mathrm{M}_z\right), c = 1, 2, \ldots, a \tag{15}$$

The local density is computed by taking the median of the η_c values, and is referred to as η_{MED}. By applying the median of the distances, the local density for each positive sample M_c via the Gaussian-like density is calculated and expressed below,

$$\Omega_c = \exp\left(-\frac{\eta_c^2}{2\eta_{MED}^2}\right) \tag{16}$$

In this case, the local density estimates for the c^{th} sample is exemplified as Ω_c, η_c represents the mean distance to its nearest neighbors, η_{MED} epitomizes the median of these average distances. Then, Ω_c is normalized as given below,

$$\Omega_c^{'} = \frac{\Omega_c}{\sum_{c=1}^{a} \alpha_c} \tag{17}$$

In the above equation, $\Omega_c^{'}$ denotes the normalized density.

The count of synthetic samples corresponding to each of the positive sample $\left\{\mathrm{M}_{c,} \mathrm{N}_c\right\}$ is first allocated according to the local density $\Omega_c^{'}$ before the samples are generated, which is stated as follows,

$$a^c = \left\lfloor \left(\left|\mathrm{M}_-\right| - \left|\mathrm{M}_+^1\right|\right) \cdot \Omega_c^{'} \right\rfloor \tag{18}$$

In the above equation, a^c typifies the synthetic samples count chosen for the positive sample.

In the process of generating samples, the process begins with identifying the k nearest neighbors of each feature vector M_c in M_+, based on the distance metric described in the above equation (14). Then, two different neighbors, M_{c_0} as well as M_{c_1}, are arbitrarily selected. Using these, a novel positive sample M_{NEW} is synthesized according to the following equation,

$$\mathrm{M}_{NEW} = \mathrm{M}_c + A_1\left(\mathrm{M}_{c_0} - \mathrm{M}_c\right) + A_2\left(\mathrm{M}_{c_1} - \mathrm{M}_c\right) \tag{19}$$

In which, A_1 and A_2 indicates the uniformly distributed random values, and W epitomizes the output of the augmented data.

3.5 Prediction of students' satisfaction

Prediction of student satisfaction involves using data analysis and modelling to anticipate how satisfied students are with their educational experience. Advantages of predicting students' satisfaction include improved teaching quality, enhanced student retention, better resource allocation, and proactive identification of academic or support issues. Existing models for predicting student satisfaction, such as standard LSTM networks, conventional DL models, or simple ML classifiers, have several limitations. They often fail to capture complex relationships between sequential VR interaction data and student profiles, lack mechanisms to prioritize the most important features, and struggle to integrate multiple sources of information effectively. Although Gaussian Process regression and other traditional machine learning models perform well on small datasets, they are less effective for sequential, high-dimensional, and multimodal data typical of VR interaction logs. These drawbacks can lead to reduced predictive accuracy, poor generalization, and limited interpretability. In this work, student satisfaction prediction is performed using a novel framework called DHA-LSTM-FH Net, which is specifically designed to effectively process metaverse-related data. The DHA-LSTM-FH Net is developed by integrating DL models, namely, the DHA-Net [40] and the LSTM model [46], employing harmonic analysis [42]. The process begins by inputting the augmented data W into both the DHA-Net method and the LSTM technique. Initially, the augmented data W is weighted using Y_1, which produces an intermediate output $\Sigma Y_1 * W$. W represents the vector, and Y_1 is a matrix, and its weighted summation produced the transformed feature vector. Subsequently, the augmented data W is also processed through the DHA-Net technique. The outcome from DHA-Net is then multiplied by the normalized result $\Sigma Y_1 * W$ to generate an intermediate outcome, indicated as I_1. Simultaneously, the same augmented data W is fed to the LSTM model, and its result is weighted employing another term, Y_2, to produce another output, I_2. Y_1 and Y_2 are learnable weight matrices, which perform linear transformations of the input features. These matrices are initialized randomly and optimized during training via backpropagation. Finally, both outputs I_1 and I_2 are fed to the DHA-LSTM-FH Net layer, where they are merged employing harmonic analysis to produce the final classification result, specified as I_3. This approach allows the model to focus on key factors influencing student satisfaction, handle complex and high-dimensional data, and produce robust and accurate predictions. Additionally, DHA-LSTM-FH Net enables better interpretability by showing which features and interactions contribute most to satisfaction outcomes, thereby supporting improved teaching quality, proactive interventions, and optimized learning experiences in Metaverse education platforms. The general outline for predicting student satisfaction with VR-based teaching in the Metaverse education platform using DHA-LSTM-FH Net is illustrated in Figure 3.

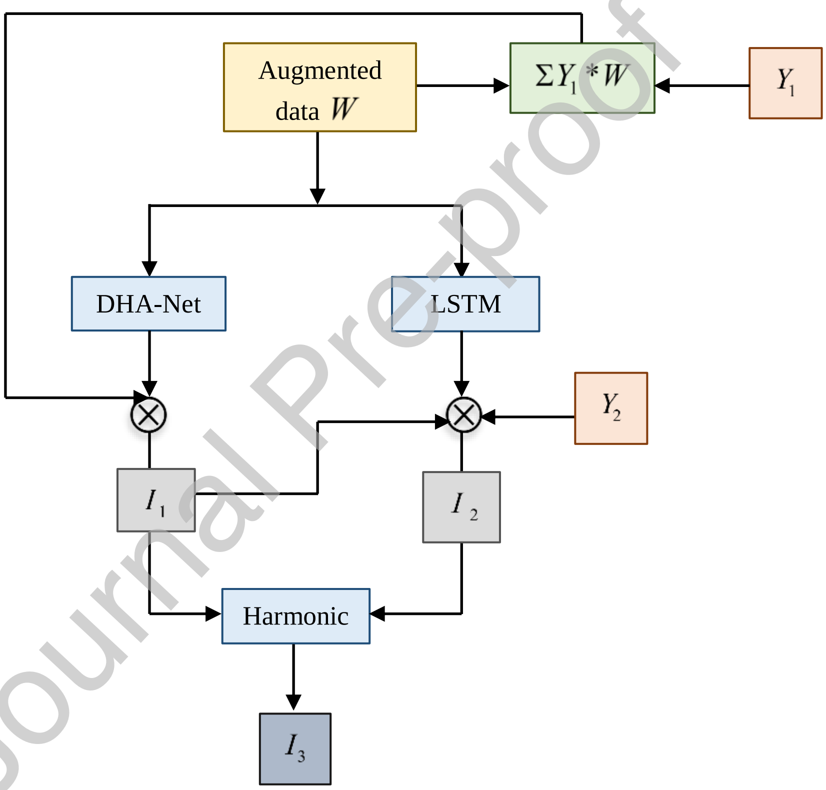


Figure 3. Block diagram of DHA-LSTM-FH Net for predicting student satisfaction with VR-based teaching in the Metaverse education platform

3.5.1 DHA-Net model

DHA-Net [40] is an advanced AI model that leverages DL with attention mechanisms to focus on key features in complex data. Conventional models may struggle to identify the most relevant behavioral and interaction patterns in VR-based learning environments, which can lower prediction accuracy. In predicting student satisfaction, DHA-Net excels by capturing nuanced emotional, behavioral, and contextual cues. As a result, the architecture provides higher prediction accuracy, better interpretability of student engagement patterns, and improved adaptability for dynamic metaverse-based learning environments. These advantages make DHA-Net a suitable component for predicting student satisfaction in VR-based teaching systems. The data augmented outcome W is provided as an input to the DHA-Net.

a) Effective Attention Module (EAM)

Attention mechanisms have recently been applied to numerous language and vision tasks, providing significant performance improvements. By integrating the Squeeze-and-Excitation Network (SENet), EAM-Net enhances this progress with a channel attention module that enables cross-channel interactions while preserving the dimensionality of the data. The process begins with global average pooling, followed by a kernel size a^1, computation and 1D convolution. Finally, a sigmoid function normalizes the output to the range (0,1). The reasons for selecting this module are outlined below.

i) No reduction in dimension

Let's represent the outcome of a basic convolutional block be epitomized as $A \in \Re^{Z \times G_2 \times G_1}$, G_1 as well as G_2 signifies the in which the height and width of the feature map, and the channel count is exemplified as Z. The weight for each channel σ_1 is computed as,

$$\sigma_1 = \nu\left(j_{\{r_1, r_2\}} e(\nu) \right) \tag{20}$$

here, the Global Average Pooling (GAP) is delineated as $e(\nu)$, and the sigmoid function is embodied as ν. The GAP layer is evaluated as,

$$e(\nu) = \frac{1}{G_2 G_1} \sum_{p=1, s=1}^{G_2 G_1} A_{:, p, s} \tag{21}$$

where, the feature map is indexed spatially, with rows and columns delineated by coordinates denoted as p and s.

Consider $\Upsilon = e(v)$, and $j_{\{r_1, r_2\}}$ is determined as,

$$j_{\{r_1, r_2\}}(\Upsilon) = u_1 J(u_{2\Upsilon}) \tag{22}$$

In this context, J specifies Rectified Linear Unit (ReLU), u_1 typifies the weights, and the input feature vector is indicated as u_2, generated by applying GAP across each channel of the feature map, is epitomized as Υ.

ii) Local cross-channel interaction

Local interactions between channels are vital for making the model more valid and efficient. The following equation demonstrates how a 1D convolution with a kernel size a^1 applied to feature Υ can easily capture interactions across channels.

$$\sigma_1 = \nu\left(\mathrm{X}_{a^1}(\Upsilon)\right) \tag{22}$$

In this context, a 1D convolution is delineated as X.

b) High-order pooling module

High-order pooling outperforms first-order pooling by capturing more comprehensive statistical information from feature representations. Recent research has demonstrated that high-order covariance pooling, when applied to first-order pooled features and normalized via matrix power, improves prediction accuracy across diverse tasks. For predicting student satisfaction with VR-based teaching in the Metaverse education platform, covariance pooling is adopted to capture second-order statistical features from the dataset, with the following sub-sections outlining the precise mathematical basis.

i) High-order Normalization

Let $\mathrm{H} \in \Re^{Z \times \vartheta \times \mathrm{E}}$ denote the input features, where the width of the input feature map is delineated as ϑ and its height is exemplified as E. The overall feature map L_1 is expressed as,

$$L_1 = \omega(\mathrm{H}; \vartheta, b_1) \tag{23}$$

wherein, ω indicates the CNN feature mapping, and b_1 exemplified the deviation of the feature map. The original feature map $L_1 \in \Re^{Z \times \vartheta \times \mathrm{E}}$ is reformatted into a feature matrix $L_1 \in \Re^{d_1 \times \mathrm{K}}$, in which $\mathrm{K} = \mathrm{E} \times \vartheta$ features, and the new number of dimensions after feature mapping is represented as d_1. The mathematical representation of the covariance matrix E_1 is formulated as,

$$E_1 = L_1 I L_1^T \tag{24}$$

where, I delineates the identity matrix, and T symbolizes the transpose of the matrix. Then, decomposing the covariance matrix E_1 via eigenvalue decomposition, it is characterized as,

$$E_1 \rightarrow (\varphi, \wedge), E_1 = \varphi \wedge \varphi^t \tag{25}$$

here, $\wedge = diagonal(y_1 \ldots y_{d_1})$ epitomizes the diagonal matrix, where $y_p (p = 1 \ldots d_1)$ represents the eigenvalue, as well as φ signifies the orthogonal matrix. The power operation of the covariance matrix E_1 is expressed by,

$$(\varphi, \wedge) \rightarrow N, N = E_1^{0.5} = \varphi \Phi(\wedge) \varphi^t \tag{26}$$

In the above equation, $\Phi(\wedge) = \sqrt{diagonal(y_1 \ldots y_{d_1})}$.

ii) Covariance Normalization

Eigenvalue decomposition plays a critical role in high-order pooling layers, but it comes with significant performance bottlenecks on GPUs due to its computational complexity. To overcome these limitations and optimize parallel computation, the Newton-Schulz Iteration (NSI) is introduced as a faster, more efficient method. NSI offers a highly efficient alternative for high-order matrix operations, replacing traditional eigenvalue decomposition, which is known for its high computational cost and inefficiency in parallel processing scenarios, and articulated as,

$$P_w = \frac{1}{2} P_{w-1}\left(3I - N_{w-1}P_{w-1}\right) \tag{27}$$

$$N_w = \frac{1}{2} N_{w-1}\left(3I - N_{w-1}P_{w-1}\right) \tag{28}$$

here, the output is signified as P. To optimize execution on GPUs, the equation is structured to support efficient matrix operations. To achieve further computational gains, the following formulation using the NSI is stated as,

$$E_1 = \frac{1}{\sqrt{tc\left(E_1\right)}} E_1 \tag{29}$$

Within the given context, $tc\left(E_1\right)$ denotes the trace of the matrix, and equation (30) can be rephrased as,

$$E_1 = \sqrt{tc\left(E_1\right) P_w} \tag{31}$$

The combination of high-order covariance pooling and EAM in DHA-Net substantially enhances student satisfaction with VR-based teaching in the Metaverse education platform. EAM focuses on extracting salient features, whereas high-order pooling captures complex inter-feature relationships. This integration leads to superior accuracy, robust generalization, and more detailed feature encoding across varied datasets. The outcome produced by employing the DHA-Net is symbolized as f_1 and is mathematically formulated by,

$$f_1 = AC\left(\xi * W + n_1\right) \tag{32}$$

In this context, the activation function is symbolized as AC, the bias is epitomized as n_1, and ξ exemplifies the weight. The EAM, inspired by SENet, applies channel-wise attention to emphasize the most relevant features from VR session data while preserving spatial dimensions. Higher-order covariance pooling then captures complex inter-feature relationships from the outputs of EAM, producing richer feature representations. This combination allows DHA-Net to focus on both salient individual features and their higher-order interactions, enhancing prediction accuracy and robustness.

The following result I_1 and is stated as,

$$I_1 = AC\left(\xi \bullet (Y_1 * W) + n_1\right) \tag{33}$$

Where, W is the input feature vector and Y_1 is a learnable transformation matrix.

The architecture of DHA-Net is demonstrated in Figure 4.

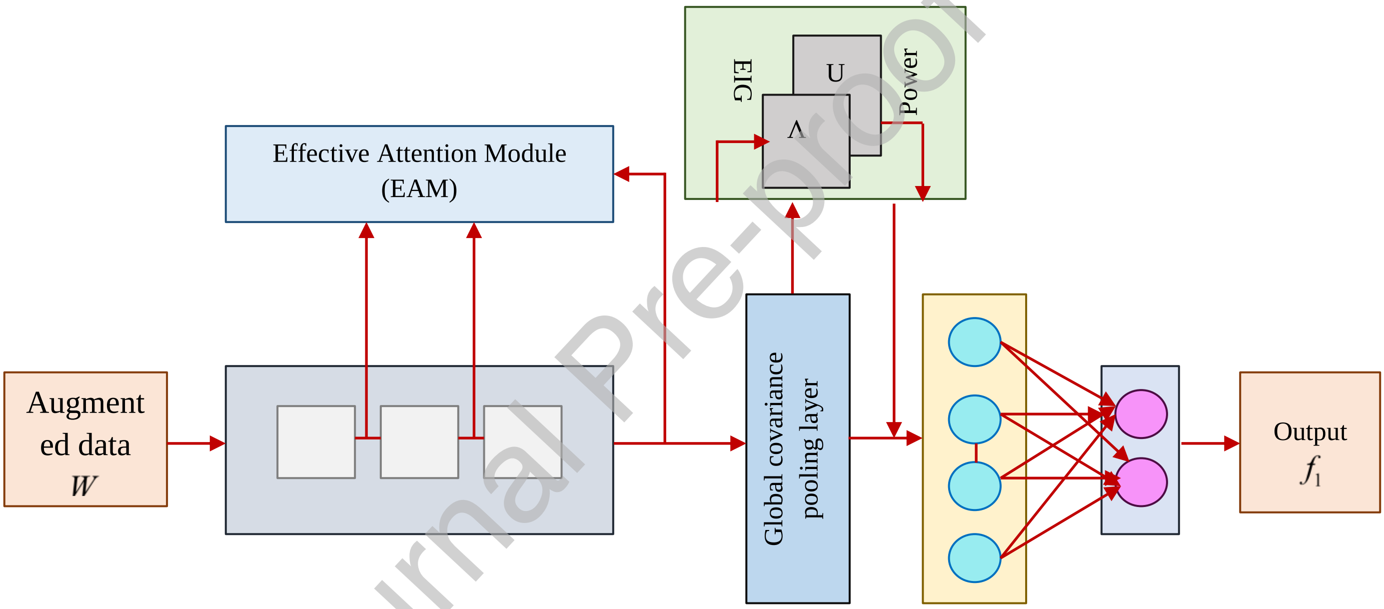


Figure 4. Structure of the DHA-Net approach

3.5.2 LSTM model

LSTM [46] extends traditional Recurrent Neural Networks (RNN) by introducing memory cells that allow them to retain important information in long-term dependencies, and effectively addresses the vanishing gradient problem. Conventional ML models and basic RNNs have been widely used for sequential data analysis; however, they often struggle to capture long-term dependencies and suffer from the vanishing gradient problem when processing large temporal datasets. Such limitations reduce their effectiveness in analyzing sequential behavioral patterns generated from VR-based learning environments. When applied to predicting students' satisfaction and satisfaction analysis, LSTMs selectively preserve the most relevant and recent academic and behavioral data, preventing memory overload and reducing noise. Consequently, the model enhances prediction accuracy, supports early identification of satisfaction trends, and enables timely interventions and personalized learning support. The augmented data W is fed as input to the LSTM technique. The cell state or memory cell, is the central element of the LSTM structure, functioning as a belt or conveyor that runs through the whole set. It allows controlled addition and removal of information, managed by gates. These gates comprise a sigmoid layer and a pointwise multiplication step, which together determine what information passes into the cell state. The gates in an LSTM are implemented using sigmoid activation functions, followed by element-wise multiplication operations. At every time step h, the LSTM cell receives the input vector $\left(S_{h-1}\right)$ and the hidden state $\left(S_h\right)$ from the previous time step, and is upgraded according to the following formulation.

The initial operation in an LSTM involves the forget gate γ_h, which evaluates and removes unnecessary information from the cell state to maintain relevant memory, and is formulated as,

$$\gamma_h = \Gamma\left(W_h E^{\gamma} + S_{h-1} \mathrm{I}^{\gamma} + g_{\gamma}\right) \tag{34}$$

where, Γ exhibits the sigmoid activation function.

Following the forget gate operation, the LSTM evaluates new information to incorporate into the cell state. This involves two distinct mechanisms, including the input gate $\left(r_h\right)$, which chooses the elements to be upgraded, and a cell state $\hat{M}_h$, which computes a set of potential values for the update. These are articulated as,

$$r_h = \Gamma\left(W_h E^{r} + S_{h-1} \mathrm{I}^{r} + g_{r}\right) \tag{35}$$

$$\hat{M}_h = \tanh\left(W_h E^M + S_{h-1} \mathrm{I}^M + g_M\right) \tag{36}$$

The updated cell state M_h is then computed by modifying the previous cell state M_{h-1}, incorporating the outputs of the forget and input gates. This update is characterized as follows,

$$M_h = M_{h-1} \otimes \gamma_h \oplus r_h \otimes \hat{M}_h \tag{37}$$

Thereafter, the output gate i_h determines which parts of the cell state contribute to the final output. After the cell state is updated, it goes through an activation function that scales its values between -1 and 1. Then, it is multiplied element-wise by the output gate to give the final hidden state, as follows,

$$i_h = \Gamma\left(W_h E^i + S_{h-1} \mathrm{I}^i + g_i\right) \tag{38}$$

$$r_1 = S_h = i_h \otimes \tanh\left(M_h\right) \tag{39}$$

In the above equation, $\mathrm{I}^\gamma, \mathrm{I}^r, \mathrm{I}^M$, and I^i epitomizes the recurrent weights, the input weights are indicated as E^γ, E^r, E^M, and E^i, and g_γ, g_r, g_M, g_i signifies the bias.

Consider, $S_h = I_2$, in equation (39)

$$I_2 = \Gamma\left(W_h E^i + S_{h-1} \mathrm{I}^i + g_i\right) \otimes \tanh\left(M_h\right) * Y_2 * I_1 \tag{40}$$

$$I_2 = \Gamma\left(W_h E^i + S_{h-1} \mathrm{I}^i + g_i\right) \otimes \tanh\left(M_h\right) * Y_2 * AC\left(\xi \bullet (Y_1 * W) + n_1\right) \tag{41}$$

The output produced from the LSTM approach is represented as I_2. The structure of the LSTM approach is illustrated in Figure 5.

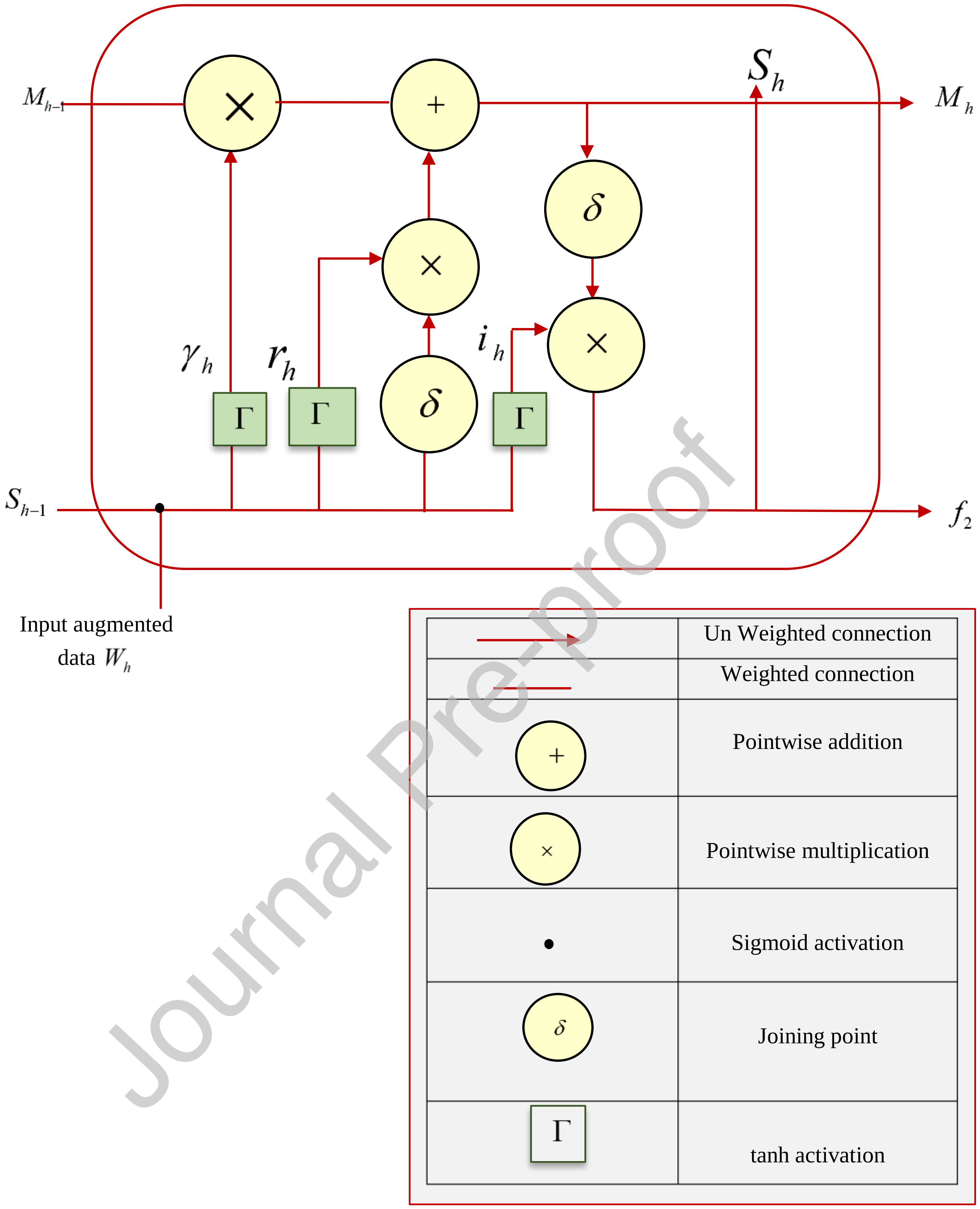


Figure 5. Structure of the LSTM approach

3.5.3 Fusion

The final result of predicting students' satisfaction in VR-based teaching within the metaverse is generated in this layer by applying harmonic analysis [42] which serves to integrate the outcomes I_1 as well as I_2. Traditional fusion methods, such as simple averaging, concatenation, stacking, or boosting, often fail to capture subtle temporal or frequency-based variations in complex student interaction data, leading to reduced accuracy and less reliable predictions. To overcome these limitations, harmonic analysis is incorporated as a mathematical approach that decomposes complex signals into basic sinusoidal waves. Unlike conventional fusion techniques, harmonic analysis transforms interaction signals into the frequency domain, enabling the model to examine periodic and latent behavioral patterns within student engagement data. This is particularly important in VR-based learning environments where student interactions occur continuously over time and often contain cyclic or repeated behavioral patterns. By doing so, it can identify subtle fluctuations, recurring patterns, and correlations in student behavior that would otherwise be missed. This enables the model to detect not only local variations, such as short-term attention shifts or interaction bursts, but also global engagement trends that remain consistent across sessions. By integrating harmonic analysis, the model is capable of detecting subtle variations in student experiences that traditional methods may overlook. This frequency-based fusion enhances the model's ability to distinguish between different types of student responses, improving classification precision. It also stabilizes predictions against temporary distractions or environmental noise, ensuring that overall patterns of engagement are accurately captured. The capacity to capture detailed frequency information from diverse behavioral patterns significantly improves the accuracy and reliability of satisfaction predictions. By analyzing harmonic components, we can distinguish more effectively between different student responses, which enhances the precision of satisfaction level classifications. Moreover, harmonic analysis aids in extracting global features of student engagement, which remain stable despite temporary distractions or local environmental variables. Overall, incorporating harmonic analysis in the fusion layer allows DHA-LSTM-FH Net to produce robust, interpretable, and highly reliable satisfaction predictions even under complex and inconsistent interaction conditions. Although the input data includes mixed feature types such as demographic information, system logs, and survey scores, each student's feature vector is treated as a pseudo-time series to enable harmonic analysis. Each element of the vector is interpreted as a sequential data point, allowing the model to capture subtle variations, correlations, and recurring patterns across features. This formulation enables the decomposition of complex, multidimensional feature interactions into harmonic components, which improves the detection of nuanced patterns related to student satisfaction. The resulting pseudo-time series representation is compatible with the DHA-LSTM-FH Net architecture, enabling the harmonic fusion layer to effectively integrate heterogeneous information.

The data is structured as a time series $m_{(1)}, m_{(2)}, \ldots m_{(R)}, \ldots m_{(\kappa)}$ of length ℓ representing the use of two consecutive time points in the harmonic analysis, based on the nature of sequential student interaction data in Metaverse VR sessions. Empirically, using two consecutive observations is sufficient to capture short-term local trends in student behavior, such as engagement spikes or drops, without introducing excessive computational complexity, where in every data point R is represented by,

$$m_{(R)} = \rho_0^{'} + \sum_{C_1=1}^{\hbar} \left(\rho_{C_1}^{'} \cos\left(\frac{2\pi C_1 R}{\ell} \right) + D_{1_\ell} \sin\left(\frac{2\pi C_1 R}{\ell} \right) \right) \tag{42}$$

Within the given context, the preselected constants are ρ and D_1, Assume $\ell = 2,\ \hbar = 1$

$$m_{(R)} = \rho_0^{'} + \rho_1^{'} \cos(\pi R) + D_{1_1} \sin(\pi R) \tag{43}$$

in which,

$$\rho_0^{'} = \frac{1}{\ell}\sum_{R=1}^{\kappa} m_R \tag{44}$$

$$\rho_0^{'} = \frac{1}{2}\left[m_{(1)} + m_{(2)}\right] \tag{45}$$

$$\rho_{C_1}^{'} = \frac{2}{\ell}\sum_{R=1}^{\kappa} m_R \cos\left(2\pi C_1 R/\ell\right) \tag{46}$$

$$\rho_1^{'} = m_{(1)}\left(-1\right) + m_{(2)}\left(1\right) \tag{47}$$

$$D_{C_1} = \frac{2}{\ell}\sum_{R=1}^{\kappa} m_R \sin\left(2\pi C_1 R/\ell\right) \tag{48}$$

$$D_1 = m_{(1)}\left(0\right) + m_{(2)}\left(0\right) \tag{49}$$

Consider the time series is modelled as $m(R-1), m(R), m(R+1)$

Let,

$$m_{(1)} = m(R-1),\ \ m_{(2)} = m(R),\ \ m_{(R)} = m(R+1) \tag{50}$$

On substituting equations (45), (46), and (50) in equation (42), we get,

$$m(R+1) = \frac{1}{2}\left[m(R-1) + m(R)\right] + \left[-m(R-1) + m(R)\right]\cos(\pi R) + 0 \tag{51}$$

$$m(R+1) = m(R-1)\left[\frac{1-2\cos(\pi R)}{2}\right] + m(R)\left[\frac{1+2\cos(\pi R)}{2}\right] \tag{52}$$

This acts as an adaptive weighting mechanism that balances the contributions of the two representations. Rather than processing the branches independently, this approach allows controlled interactions between them, capturing both complementary and contrasting relationships among features.

Consider, $m(R-1) = I_1$, $m(R) = I_2$, $m(R+1) = I_3$

$$I_3 = I_1\left[\frac{1-2\cos(\pi R)}{2}\right] + I_2\left[\frac{1+2\cos(\pi R)}{2}\right] \tag{53}$$

$$I_3 = AC\left(\xi \bullet (Y_1 * W) + n_1\right)\left[\frac{1-2\cos\left(\pi R\right)}{2}\right] + \left(\begin{array}{l}\Gamma\left(W_h E^i + S_{h-1} \mathrm{I}^i + g_i\right) \otimes \tanh\left(M_h\right) * Y_2 \\ * AC\left(\xi \bullet (Y_1 * W) + n_1\right)\left[\frac{1+2\cos\left(\pi R\right)}{2}\right]\end{array}\right) \tag{54}$$

wherein, I_3 symbolizes the result of the DHA-Net model, which demonstrates the predicted student's satisfaction outcome. $\cos\left(\pi R\right)$ serve as modulation coefficients that dynamically regulate the contribution of different feature components. These coefficients are based on harmonic weighting principles, using cosine waves to capture both constructive and destructive interactions between signals.

The developed DHA-LSTM-FH Net's pseudocode is illustrated in Algorithm 2.

Algorithm 2. Pseudocode of the developed DHA-LSTM-FH Net

Input	
	Training: Augmented data W and Training labels
	Testing: Testing data W_{test} , Y_1, Y_2, Weight of LSTM, Weight of DHA-Net,
Output	
	Weights of DHA-LSTM-FH Net
	Predicted labels of DHA-LSTM-FH Net
Parameters required	
	Learning rate A_2
	Batch size A_3
	Maximum Epoch A_1
Begin	
Initialization	
Randomly initialize Y_1, Y_2 Weight of the DHA-Net and the weight of the LSTM	
Set the learning rate, batch size, and epoch	
//Training stage	
	while $h_1 < \mathrm{A}_1$
	for every augmented data W
	Multiplying the augmented data W and the weight Y_1, yielding the desired result $\Sigma Y_1 * W$
	Execute DHA-Net and multiply it to get I_1 $I_1 = AC\left(\xi \bullet (Y_1 * W) + n_1\right)$
	The augmented data W is fed to the LSTM model, and its output is weighted by term, Y_2, to produce another output, I_2 . $I_2 = \Gamma\left(W_h E^i + S_{h-1} \mathrm{I}^i + g_i\right) \otimes \tanh\left(M_h\right) * Y_2 * AC\left(\xi \bullet (Y_1 * W) + n_1\right)$

	To obtain the result I_3, perform harmonic analysis by merging I_1 and I_2. $$I_3 = AC\left(\xi \bullet (Y_1 * W) + n_1\right)\left[\frac{1-2\cos(\pi R)}{2}\right] + \left(\begin{array}{l}\Gamma\left(W_h E^i + S_{h-1} I^i + g_i\right) \otimes \tanh\left(M_h\right) * Y_2 \\ * AC\left(\xi \bullet (Y_1 * W) + n_1\right)\left[\frac{1+2\cos(\pi R)}{2}\right]\end{array}\right)$$
	end for
	The error E_r is calculated by considering the training labels and I_3
	The weights and biases are updated by DHA-LSTM-FH Net using the Adam optimizer with MSE loss.
	End while
	Save the optimal Weight of LSTM, weight of DHA-Net, Y_1 and Y_2
// testing stage	
	For each test, augmented data W
	Determine I_3, I_1 and I_2 utilizing trained weights of DHA-LSTM-FH Net (Weight of LSTM, weight of DHA-Net, Y_1 and Y_2)
	Attain the predicted label I_3
	End for
End	

4. Result and discussion

The DHA-LSTM-FH Net introduced in this research is evaluated for its ability to predict student satisfaction in this section, emphasizing various performance metrics and comparisons with existing methods.

4.1 Experimental set-up

The proposed system is implemented and evaluated using a system configured with Windows 11 as the operating system. The hardware environment includes 8 GB of RAM and more than 100 GB of ROM to support data storage and processing requirements. The system is equipped with a CPU operating at 1.7 GHz and supports GPU acceleration, which improves the efficiency of DL computations and model training. For the software environment, the implementation is carried out using the Python programming language, version 3.9.11, which provides extensive libraries and frameworks for ML and DL development.

4.2 Dataset description

Impact of Virtual Reality on Education [43] is a comprehensive dataset that examines the role of VR technology in enhancing educational experiences across multiple disciplines. It is designed to facilitate the analysis of the effectiveness of immersive learning compared to traditional pedagogical methods. The dataset comprises both quantitative and qualitative data collected from diverse educational institutions that have integrated VR into their teaching practices. The dataset includes student demographic profiles, VR usage statistics, and feedback from both students and educators. In addition, it provides several educational impact metrics, such as student engagement scores, retention rates, academic performance grades, session duration, and immersion levels experienced during VR-based learning sessions. The dataset also contains teacher and student feedback scores that reflect perceptions regarding the effectiveness and engagement level of VR-assisted instruction across different academic subjects. This resource is particularly valuable for researchers, educators, and policymakers seeking to evaluate the impact of VR in academic environments. It has a dimension of 5000×20 ,where 5000 represents the number of samples, and 20 represents the number of features describing student interactions, engagement indicators, and learning activities in the virtual learning environment.

The VR_College_Teaching_Dataset [44] is designed to model a VR teaching scenario in a college context. It captures comprehensive information, including student demographics, VR interaction logs, engagement levels, academic performance, and technical infrastructure data. The dataset includes attributes related to student information, such as course, year of study, and prior VR experience. It also records VR session interaction data, including engagement time, task completion rate, and quiz scores. Additionally, performance metrics such as knowledge retention, skill acquisition score, and user feedback are provided to evaluate learning outcomes. Technical parameters like VR module type, system usage, lag time, and connection stability, along with infrastructure factors, such as classroom type, teacher training level, and technical issues faced, are also included. Outcome metrics, such as satisfaction score, future interest, and comparison with traditional teaching methods, are provided, where the comparison variable serves as the target attribute with three categories: Better, Neutral, and Worse. The primary objective of the dataset is to assess and predict student satisfaction with VR-based instruction relative to conventional teaching methods. It has a dimension of 200×23, where 200 denotes the number of samples, and 23 denotes the number of features related to VR-based teaching sessions and student behavioral responses.

The dataset is partitioned using a random splitting strategy to ensure unbiased model training and evaluation. Initially, the dataset is randomly shuffled and then divided into three subsets: 90% of the data is used for training, 5% for testing, and the remaining 5% for validation. This random partitioning enables the proposed model to be trained effectively while ensuring reliable evaluation and parameter tuning through independent testing and validation sets. In the following sections, the impact of Virtual Reality on the Education Dataset is illustrated as Dataset 1, and the VR_College_Teaching_Dataset is depicted as Dataset 2. These two datasets are utilized in this study with different target representations. Dataset 1 contains binary labels representing the comparison between VR-based learning and traditional teaching methods, allowing direct evaluation using classification metrics such as precision, recall, and F1-score. In contrast, Dataset 2 provides satisfaction scores obtained from survey responses. To enable classification-based evaluation, these continuous satisfaction scores are categorized into discrete satisfaction levels using predefined threshold ranges before computing the performance metrics.

4.3 Experimental Parameters

The experimental parameters of DHA-LSTM-FH Net are depicted in Table 1.

Table 1. The experimental parameters of DHA-LSTM-FH Net

Parameters	Values
Optimizer	Adam
Dropout	0.2
Batch size	32
Epoch	80
Activation	ReLU
Loss	MSE
Kernel size	(3,3)
Learning rate	0.01
Beta 2	0.9999
Epsilon	0.00001
Beta 1	0.9
Network dimensions	4
Attention heads	4
LSTM units	128

4.4 Performance metrics

The effectiveness of the DHA-LSTM-FH Net is analyzed employing performance metrics, like precision, recall, F1 Score, and Cohen's Kappa.

i) Precision

Precision (T_{PRE}) in student satisfaction prediction measures the accuracy of positive predictions made by the model, that is, the proportion of students predicted to be satisfied who are truly satisfied. Mathematically, precision is defined as,

$$\mathrm{T}_{PRE} = \frac{O_1}{O_1 + O_2} \tag{55}$$

wherein, O_1 is epitomized as the true positive, and the false positive is specified as O_2.

ii) Recall

Recall (T_{RE}) measures the model's ability to identify all truly satisfied students. It is the proportion of actually satisfied students who are correctly predicted as satisfied by the model. Recall is mathematically expressed as,

$$\mathrm{T}_{RE} = \frac{O_1}{O_1 + O_3} \tag{56}$$

where, the false negative is represented as O_3.

iii) F1 Score

The F1 Score (T_{SC}) is the harmonic mean of precision and recall, offering a single metric that balances the trade-off between false positives and false negatives. In the context of student satisfaction prediction, the F1 Score reflects the model's overall effectiveness in both correctly identifying satisfied students (precision) and capturing all truly satisfied students (recall). It is calculated as,

$$\mathrm{T}_{SC} = 2 \times \frac{\mathrm{T}_{PRE} \times \mathrm{T}_{RE}}{\mathrm{T}_{PRE} + \mathrm{T}_{RE}} \tag{57}$$

iv) Cohen's Kappa (κ)

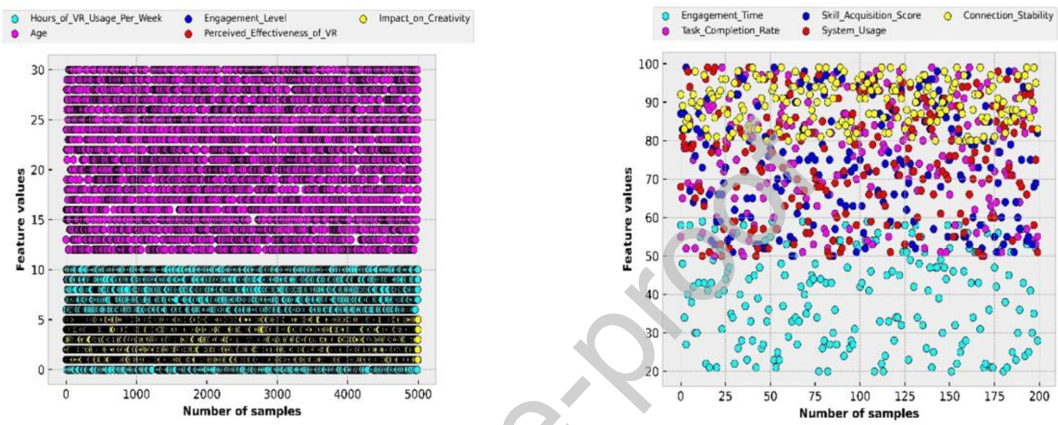
It is a statistical measure used to evaluate the level of agreement between two raters or classification methods when assigning items into categorical classes. Unlike simple accuracy, Cohen's Kappa takes into account the agreement that may occur by chance, providing a more reliable measure of classification performance.

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{58}$$

Where, P_o is the observed agreement, and P_e represents the expected agreement by chance.

4.5 Data Visualization

The data visualization of both Datasets is presented in Figure 6. Figure 6(a) illustrates several features from the Impact of Virtual Reality on Education dataset, including weekly VR usage hours, impact on creativity, engagement level, perceived effectiveness of VR, and age. The yaxis represents the feature values, while the x-axis shows the number of samples. The age feature ranges from 12 to 30. The impact on the creativity feature ranges from 0 to 5, and the hours of VR usage per week range from 0 to 10. In Figure 6(b), features from the VR_College_Teaching_Dataset, including engagement time, skill acquisition score, connection stability, task completion rate, and system usage, are demonstrated. Engagement time ranges from 20 to 60, while connection stability ranges from 80 to 100.



(a) **(b)**

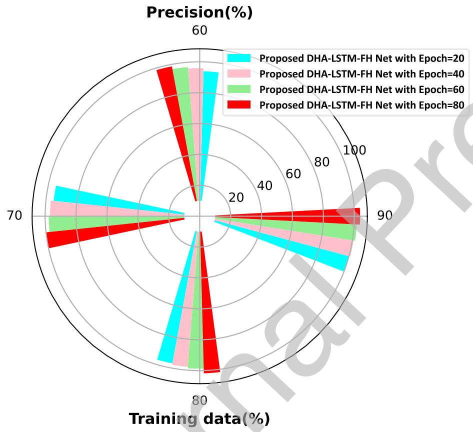
Figure 6. Data visualization of a) Dataset 1 and b) Dataset 2

4.6 Performance assessment

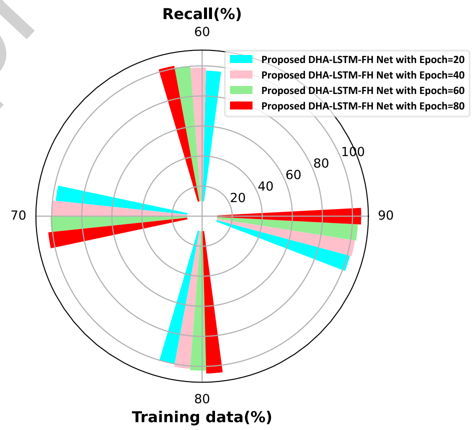
This section presents the assessment of the DHA-LSTM-FH Net in terms of predicting student satisfaction, focusing on its performance metrics across varying epochs and training data for both Datasets.

4.6.1 Analysis concerning Dataset 1

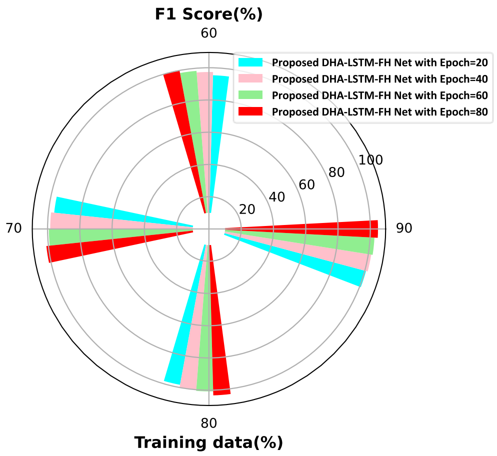
The performance evaluation of the DHA-LSTM-FH Net across different epochs and training data in Dataset 1 is illustrated in Figure 7. The colored bars represent the performance of the proposed DHA-LSTM-FH Net model at different training epochs. Specifically, the red bar indicates the model trained with Epoch 20, the cyan (light blue) bar represents Epoch 40, the green bar corresponds to Epoch 60, and the red bar denotes Epoch 80. These colored bars are used to illustrate the variation in performance of the proposed model as the number of training epochs increases, showing how the model's prediction capability improves with extended training. In Figure 7(a), the performance of DHA-LSTM-FH Net evaluated with respect to precision is depicted. The precision attained by DHA-LSTM-FH Net is 86.876%, 87.876%, 88.765%, and 91.765% for 90% of the training data and 20, 40, 60, and 80 epochs. The recall-based investigation of DHA-LSTM-FH Net is depicted in Figure 7(b). For 20, 40, 60, and 80 epochs, the recall gained by the DHA-LSTM-FH Net model is 87.877%, 89.876%, 90.765%, and 92.865% for training data 70%. The analysis of DHA-LSTM-FH Net in terms of F1 Score is demonstrated in Figure 7(c). The DHA-LSTM-FH Net obtained an F1 Score of 90.754%, 91.808%, 92.700%, and 94.750% for 20, 40, 60, and 80 epochs, and 90% of the training data.



(a)



(b)

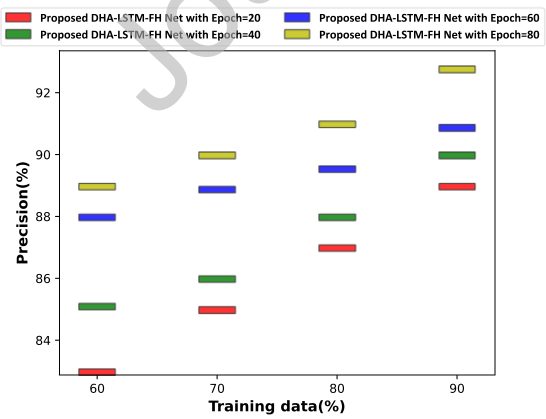


(c)

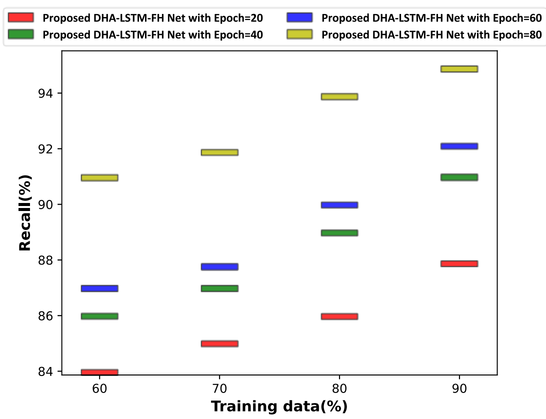
Figure 7. Performance assessment of DHA-LSTM-FH Net for predicting students' satisfaction in Dataset 1, regarding a) Precision, b) Recall, and c) F1 Score

4.6.2 Evaluation in terms of Dataset 2

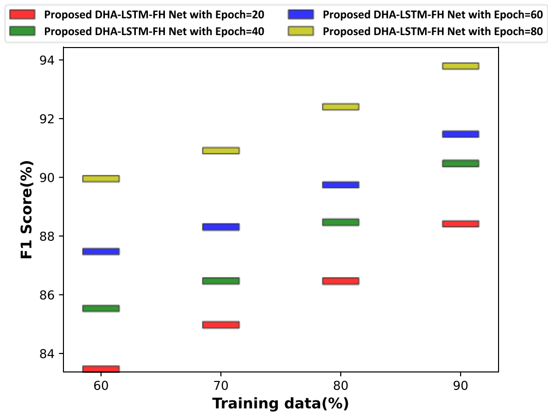
In Figure 8, the performance analysis of the DHA-LSTM-FH Network, evaluated across various epochs and training data within Dataset 2, is demonstrated. The colored markers represent the performance of the proposed DHA-LSTM-FH Net model at different training epochs. The red marker indicates the model trained with Epoch 20, the green marker represents Epoch 40, the blue marker corresponds to Epoch 60, and the yellow marker denotes Epoch 80. These colored markers illustrate the variation in the proposed model's performance across different training data percentages and epochs, showing that performance gradually improves as both the training data and the number of epochs increase. Figure 8 (a) illustrates the analysis of DHA-LSTM-FH Net using precision. For 20, 40, 60, and 80 epochs, and 90% of the training data, the DHA-LSTM-FH Net achieved a precision of 88.87%, 89.876%, 90.765%, and 92.654%, respectively. Figure 8(b) illustrates the recall analysis of the DHA-LSTM-FH Net approach. For 60% of the training data, the recall obtained by DHA-LSTM-FH Net is 83.865%, 85.876%, 86.876%, and 90.866% for 20, 40, 60, and 80 epochs. Figure 8(c) illustrates the performance of DSNA-Net in terms of F1 Score. The F1 Score obtained by the DHA-LSTM-FH Net approach is 83.368%, 85.429%, 87.373%, and 89.860% for 60% of the training data and 20, 40, 60, and 80 epochs, respectively.



(a)



(b)



(c)

Figure 8. Performance valuation of DHA-LSTM-FH Net for predicting students' satisfaction using Dataset 2 in terms of a) Precision, b) Recall, and c) F1 Score

4.7 Comparative techniques

The effectiveness of DHA-LSTM-FH Net for predicting student satisfaction is evaluated in this section through a comparison with prevailing models, such as Hybrid CB-SEM and Deep ANN Models [6], PLS-SEM combined with ANN [29], Hybrid SEM-ML [30], TAM [33], ANN [35], E-SATNet [35], and XGBoost [36].

4.8 Comparative evaluation

This section provides a detailed comparison of DHA-LSTM-FH Net's performance across different training data for Datasets 1 and 2.

4.8.1 Comparative based on Dataset 1

Figure 9 illustrates the analysis of the proposed DHA-LSTM-FH Net in predicting student satisfaction using Dataset 1. The differently colored lines represent various existing prediction models compared with the proposed DHA-LSTM-FH Net method. The blue line indicates the Hybrid CSB-SEM and Deep ANN model, the green line represents the PLS-SEM and ANN models, the orange line corresponds to the Hybrid SEM-ML model, the magenta line denotes the ANN model, the yellow line represents the TAM model, the grey line indicates the E-SAT-Net model, and the brown line shows the XGBoost technique. The red star line represents the proposed DHA-LSTM-FH Net model. The analysis of the DHA-LSTM-FH Net in terms of precision is shown in Figure 9(a). With 80% of the training data, DHA-LSTM-FH Net reached a precision of 91.765%, whereas the current techniques, like hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost obtained a precision of 82.877%, 86.765%, 87.877%, 88.389%, 88.876%, 89.478%, and 90.666%. Notably, the DHA-LSTM-FH Net shows a 4.24% precision improvement over the hybrid SEM-ML model. In Figure 9 (b), the examination of DHA-LSTM-FH Net based on recall is epitomized. Prevailing models, like hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, reached a recall of 87.876%, 91.876%, 92.876%, 93.008%, 93.544%, 94.368%, and 94.777%, for 90% of the training data, whereas the proposed model obtained a recall of 95.755% for 90% of the training data. In this context, DHA-LSTM-FH Net outperforms PLS-SEM and ANN model with a 4.05% improvement in recall. The evaluation of the developed DHA-LSTM-FH Net concerning the F1 Score is indicated in Figure 9 (c). The proposed DHA-LSTM-FH Net model achieved an F1 Score of 91.289%, whereas the existing models, such as hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, obtained an F1 Score of 82.813%, 85.350%, 87.345%, 87.846%, 88.245%, 88.843%, and 90.025%, for 70% of the training data. DHA-LSTM-FH Net achieves a 4.32% higher F1 Score compared to the hybrid SEM-ML model. The Cohen's Kappa evaluation is illustrated in Figure 9(d). For 60% of the training data, the existing approaches, including hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, recorded Cohen's Kappa values of 0.689, 0.719, 0.739, 0.749, 0.759, 0.766, and 0.778, respectively. In comparison, the proposed DHA-LSTM-FH Net produced a Cohen's Kappa value of 0.789, indicating better agreement between predicted and actual classifications. Furthermore, the proposed DHA-LSTM-FH Net demonstrates a 2.92% improvement in Cohen's Kappa compared to the E-SATNet model.

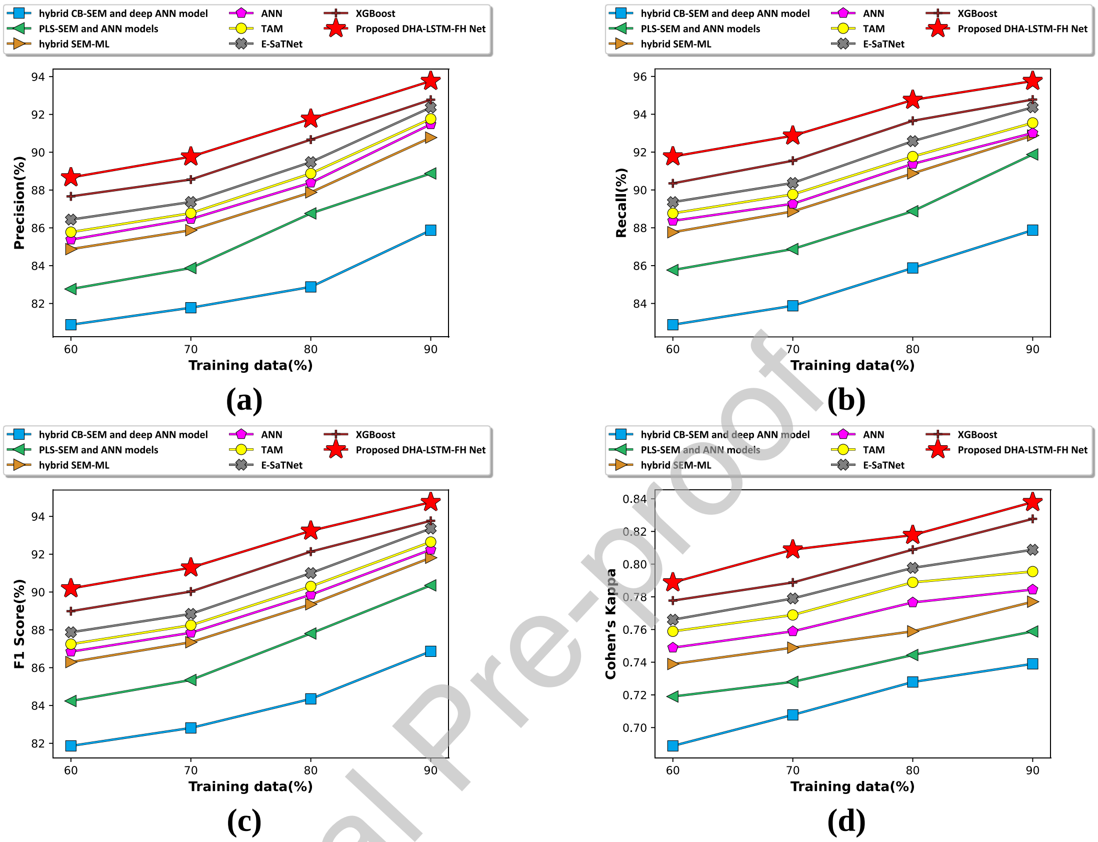


Figure 9. Comparative analysis of DHA-LSTM-FH Net for forecasting student satisfaction based on a) Precision, b) Recall, c) F1 Score, and d) Cohen's Kappa

4.8.2 Comparative assessment with respect to Dataset 2

The effectiveness of the DHA-LSTM-FH Net model in forecasting student satisfaction, based on Dataset 2, is shown in Figure 10. The colored bars represent different prediction models used for comparative evaluation. The grey bar denotes the Hybrid CSB-SEM and Deep ANN model, the brown bar represents the PLS-SEM and ANN models, the gold bar corresponds to the Hybrid SEM-ML model, the green bar indicates the ANN model, the pink bar represents the TAM model, the dark blue bar denotes the E-SAT-Net model, and the orange bar depicts the XGBoost model. The red bar represents the proposed DHA-LSTM-FH Net model. In Figure 10(a), the investigation of DHA-LSTM-FH Net based on precision is illustrated. For 90% of the training data, the proposed DHA-LSTM-FH Net model achieved a precision of 92.654%. In comparison, existing models such as the hybrid CB-SEM and deep ANN model, PLS-SEM and ANN models, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, reached precisions of 86.876%, 87.877%, 88.765%, 89.257%, 89.765%, 90.478%, and 91.777%, respectively. Compared to the TAM model, the DHA-LSTM-FH Net demonstrates a 3.12% improvement in precision. The recall evaluation of DHA-LSTM-FH Net is shown in Figure 10(b). For 90% of the training data, prevailing models like hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, achieved recalls of 86.790%, 88.877%, 91.876%, 92.368%, 92.665%, 93.467%, and 93.889%, respectively. On the other hand, the proposed DHA-LSTM-FH Net attained a recall of 94.765%. The DHA-LSTM-FH Net shows a 2.22% increase in the recall compared to the TAM. Figure 10(c) presents the evaluation of the proposed DHA-LSTM-FH Net in terms of F1 Score. The developed model achieved an F1 Score of 92.303% using 80% of the training data. In comparison, existing models such as hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, attained F1 Score of 84.298%, 85.865%, 88.379%, 88.877%, 89.317%, 90.066%, and 91.194%, respectively. The DHA-LSTM-FH Net demonstrates a 2.67% improvement over the TAM in F1 Score. In Figure 10(d), the analysis of Cohen's Kappa is presented. The proposed DHA-LSTM-FH Net approach records the highest Cohen's Kappa value of 0.759, whereas the traditional techniques, including hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, report Cohen's Kappa values of 0.659, 0.670, 0.699, 0.709, 0.720, 0.739, and 0.748, respectively, for 70% of the training data. Compared to the XGBoost model, the DHA-LSTM-FH Net shows a 2.69% improvement in Cohen's Kappa, indicating stronger agreement between predicted and actual classifications.

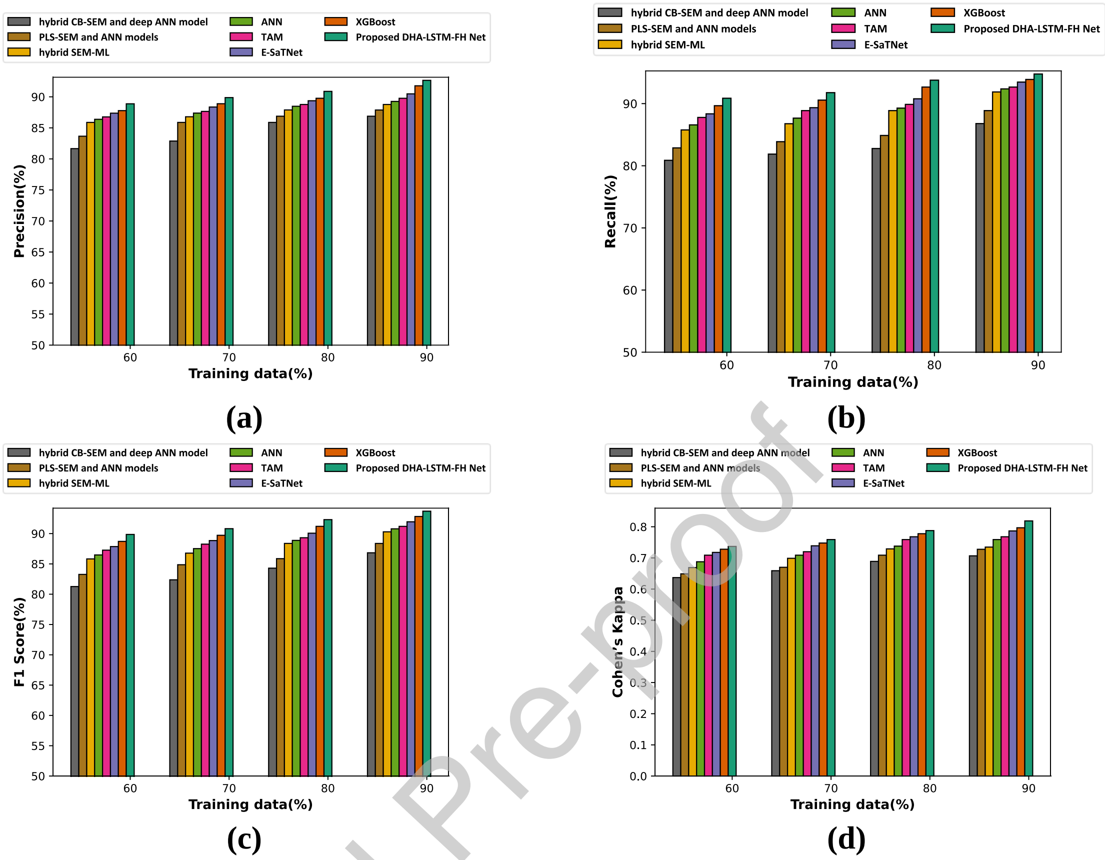
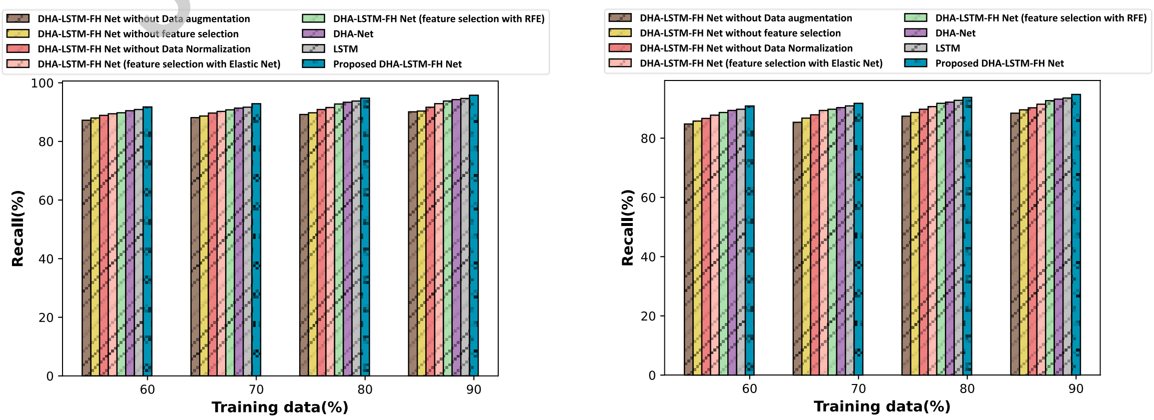


Figure 10. Comparative evaluation of DHA-LSTM-FH Net for predicting student satisfaction concerning a) Precision, b) Recall, c) F1 Score, and d) Cohen's Kappa

4.9 Ablation analysis

Ablation analysis is a technique used in ML and DL to evaluate the contribution of individual components or modules of a model. In this process, specific parts of the model are systematically removed or modified, and the model's performance is measured without them. By comparing the results with the full model, researchers can identify which components are most critical for overall performance and understand the impact of each module on predictive accuracy, robustness, or efficiency. Figure 11 illustrates the ablation study analysis, showing the variation of recall with different amounts of training data for both datasets. The performance of the proposed DHA-LSTM-FH Net is compared with its individual components and variations, including: DHA-LSTM-FH Net without data normalization, DHA-LSTM-FH Net with feature selection using Elastic Net, DHA-LSTM-FH Net with feature selection using RFE, DHA-LSTM-FH Net without feature selection, DHA-LSTM-FH Net without data augmentation, DHA-Net, and LSTM. In Figure 11 (a), the ablation analysis for Dataset 1 is presented. For 90% of the training data, the proposed DHA-LSTM-FH Net records the highest recall of 95.755%, while the individual modules report recall values of 90.098%, 90.368%, 91.666%, 92.888%, 93.777%, 94.269%, and 94.668%, respectively. Figure 11 (b) illustrates the ablation study for Dataset 2, where the discrete components achieve recall values of 88.437%, 89.567%, 90.257%, 91.467%, 92.666%, 93.178%, and 93.555%, whereas the proposed DHA-LSTM-FH Net model attains the highest recall of 94.765% for 90% of the training data. These results indicate that each component of the DHA-LSTM-FH Net contributes positively to the overall performance, and the combination of feature selection, data normalization, data augmentation, DHA-Net and LSTM significantly enhances recall. The ablation analysis highlights the robustness and effectiveness of the integrated framework, demonstrating that the proposed model consistently outperforms its individual modules across both datasets.

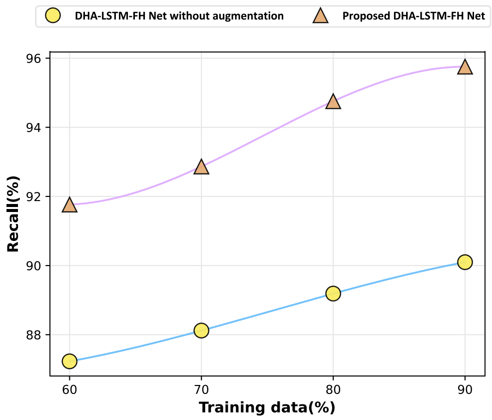


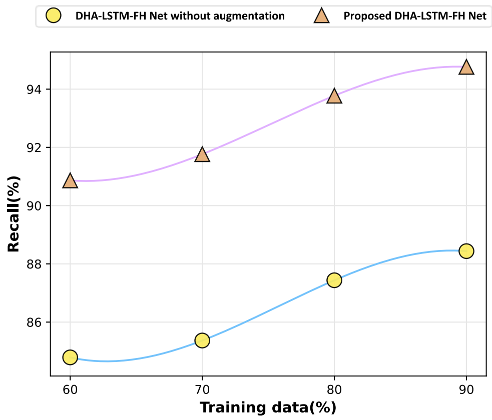
(a) **(b)**

Figure 11: Ablation study for a) Dataset 1 and b) Dataset 2

4.10 Augmentation analysis

Augmentation analysis is a technique used to evaluate the impact of data augmentation methods on the performance of an ML model. It involves systematically comparing the model's performance with and without augmentation to determine how generating additional synthetic data improves metrics such as recall. This analysis helps demonstrate the effectiveness of augmentation techniques in enhancing model generalization, reducing overfitting, and improving robustness, especially for imbalanced datasets. Figure 12 illustrates the augmentation analysis for both datasets, showing the variation in recall with different proportions of training data. The performance of the DHA-LSTM-FH Net without data augmentation is compared to the proposed DHA-LSTM-FH Net with augmentation. In Figure 12 (a), the augmentation analysis for Dataset 1 is presented. For 90% of the training data, the model without data augmentation records a recall of 90.098%, whereas the developed DHA-LSTM-FH Net reaches a recall of 95.755%. Similarly, Figure 12 (b) illustrates the augmentation analysis for Dataset 2, where the model without augmentation achieves a recall of 88.437%, while the proposed DHA-LSTM-FH Net records a recall of 94.765% for the same training data proportion. These results demonstrate that data augmentation significantly improves the model's ability to generalize, particularly for complex and imbalanced Metaverse interaction data. By generating additional synthetic samples, the proposed DHA-LSTM-FH Net becomes more robust, reduces overfitting, and consistently enhances recall across both datasets, confirming the effectiveness of the augmentation strategy in improving overall predictive performance.



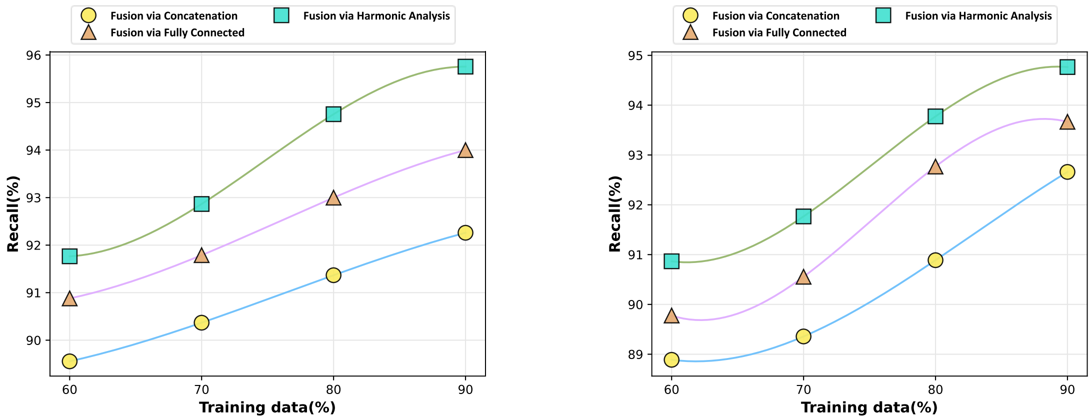


(a) **(b)**

Figure 12: Augmentation analysis for a) Dataset 1 and b) Dataset 2

4.11 Fusion analysis

Fusion analysis is a technique used to evaluate the effectiveness of different methods for combining multiple model outputs or feature representations. In Figure 13, the fusion analysis for both datasets is presented, showing the variation of recall with different proportions of training data. The performance of Fusion via Concatenation, Fusion via Fully Connected layers, and Fusion via Harmonic Analysis is compared. Figure 13 (a) depicts the results for Dataset 1, where, for 90% of the training data, the concatenation and fully connected fusion models achieve recall values of 92.260% and 93.999%, respectively, while harmonic analysis records the highest recall of 95.755%. Similarly, Figure 13 (b) illustrates the fusion analysis for Dataset 2, where the concatenation and fully connected methods obtain recall values of 92.660% and 93.666%, whereas harmonic analysis reaches a recall of 94.765% for 90% of the training data. These results indicate that harmonic analysis provides the most effective fusion strategy, as it captures both local and global patterns in student interaction data more comprehensively than simple concatenation or fully connected fusion. By integrating frequency-based information, the model enhances predictive accuracy, improves robustness to variations in student behavior, and ensures consistent performance across diverse datasets.



(a) **(b)**

Figure 13: Fusion analysis for a) Dataset 1 and b) Dataset 2

4.12 Statistical test analysis

To determine whether the improvements of the proposed approach are statistically significant compared with baseline models, paired t-tests are performed on several evaluation metrics. The analysis uses the results obtained from K-fold cross-validation with K =10 on both datasets. For each metric, the fold-wise performance of the proposed method is compared with that of each baseline model, and a p-value below 0.05 is considered statistically significant. In addition, the statistical significance of the proposed DHA-LSTM-FH Net method is further examined using paired t-tests on precision, recall, F1 Score, and Cohen's Kappa across 5-fold cross-validation for both datasets. The detailed results of these tests are summarized in Table 2. The analysis indicates that the proposed approach consistently performs better than the baseline methods, including hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, across all four metrics for both datasets. For Dataset-1, the comparisons produce statistically significant results ($p < 0.05$) in all cases. The proposed method demonstrates higher precision than hybrid CB-SEM and deep ANN ($t = 3.28$, $p = 0.02$), PLS-SEM and ANN ($t = 2.88$, $p = 0.02$), hybrid SEM-ML ($t = 2.59$, $p = 0.02$), ANN ($t = 2.37$, $p = 0.03$), TAM ($t = 2.12$, $p = 0.03$), E-SATNet ($t = 1.89$, $p = 0.03$), and XGBoost ($t = 1.70$, $p = 0.03$). Similar statistically significant improvements are observed for recall, F1

Score, and Cohen's Kappa. The findings remain consistent for Dataset-2, where the proposed approach again demonstrates superior performance across all metrics. For example, when compared with hybrid CB-SEM and deep ANN, the t-statistic values are 3.49, 3.69, 3.31, and 3.10 for precision, recall, F1 Score, and Cohen's Kappa, respectively, with p-values of 0.01 and 0.02. Likewise, significant differences are also observed against PLS-SEM and ANN (p = 0.02), hybrid SEM-ML (p = 0.02 and 0.03), ANN (p = 0.02 and 0.03), TAM (p = 0.02 and 0.03), E-SATNet (p = 0.03 and 0.04), and XGBoost (p = 0.03 and 0.04). Overall, the statistically significant improvements obtained across both datasets confirm the robustness and effectiveness of the proposed method, demonstrating its ability to deliver superior classification performance compared with conventional baseline approaches.

Table 2: T-test analysis

Proposed	Comparative methods	T-statistics				P-value			
		Precision	Recall	F1 Score	Cohen's Kappa	Precision	Recall	F1 Score	Cohen's Kappa
Dataset 1									
Proposed DHA-LSTM-FH Net	**Hybrid CB-SEM and deep ANN model**	3.28	3.48	3.10	2.89	0.02	0.01	0.02	0.02
	PLS-SEM and ANN models	2.88	3.38	2.80	2.59	0.02	0.02	0.02	0.02
	hybrid SEM-ML	2.59	3.10	2.49	2.22	0.02	0.02	0.02	0.03
	ANN	2.37	2.80	2.12	1.99	0.03	0.02	0.03	0.03
	TAM	2.12	2.59	1.89	1.80	0.03	0.03	0.03	0.03
	E-SaTNet	1.89	2.38	1.59	1.49	0.03	0.03	0.04	0.04
	XGBoost	1.70	2.18	1.38	1.23	0.03	0.03	0.04	0.04
Dataset 2									
Proposed DHA-LSTM-FH Net	**Hybrid CB-SEM and deep ANN model**	3.49	3.69	3.31	3.10	0.01	0.01	0.02	0.02
	PLS-SEM and ANN models	3.09	3.59	3.01	2.80	0.02	0.02	0.02	0.02
	hybrid SEM-ML	2.80	3.31	2.70	2.44	0.02	0.02	0.02	0.03
	ANN	2.58	3.01	2.33	2.20	0.02	0.02	0.03	0.03
	TAM	2.33	2.80	2.10	2.01	0.03	0.02	0.03	0.03
	E-SaTNet	2.10	2.59	1.80	1.70	0.03	0.03	0.03	0.04
	XGBoost	1.91	2.39	1.59	1.44	0.03	0.03	0.04	0.04

4.13 Confusion matrix

A confusion matrix is a performance evaluation tool used in classification problems to measure how well a machine learning model predicts classes. It compares the actual class labels with the predicted class labels produced by the model. The matrix typically contains four values: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These values help evaluate metrics such as precision, recall, F1 Score, and Cohen's Kappa, providing a detailed understanding of the model's prediction performance. For Dataset 1, Figure 14 illustrates the confusion matrix for the proposed DHA-LSTM-FH Net model in predicting student satisfaction in the Metaverse learning platform with two classes: Satisfied and Not Satisfied. Out of the total predictions, 288 instances are correctly classified as Not satisfied, while 276 instances are correctly classified as Satisfied. The model incorrectly predicts 20 instances as Satisfied when they are actually Not satisfied, and 19 instances are predicted as Not satisfied when they are actually Satisfied. These results indicate that the model correctly identifies most student satisfaction levels with only a small number of misclassifications. Overall, the confusion matrix demonstrates that the proposed DHA-LSTM-FH Net model achieves high prediction accuracy and balanced classification performance for both satisfied and not satisfied students. This indicates that the model is effective in analyzing student satisfaction in Metaverse-based learning environments and supports reliable educational decision-making.

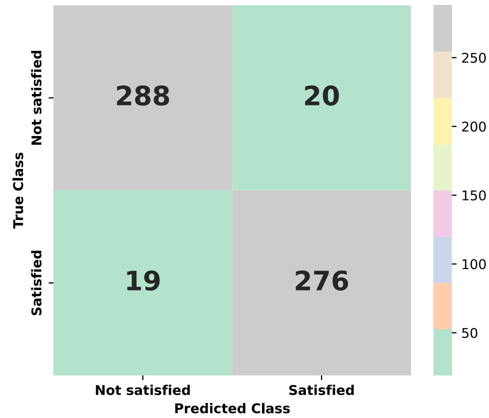


Figure 14: Confusion matrix for Dataset 1

4.14 Comparative discussion

The comparative analysis of DHA-LSTM-FH Net for predicting student satisfaction based on both Datasets with 90% of the training data is depicted in Table 3. The DHA-LSTM-FH Net attained a precision of 93.765%, a recall of 95.755%, an F1 Score of 94.750%, and a Cohen's Kappa of 0.838 for Dataset 1 with a training data of 90%. The prevailing models, like hybrid CB-SEM and deep ANN, PLS-SEM and ANN, hybrid SEM-ML, ANN, TAM, E-SATNet, and XGBoost, obtained a precision of 87.876%, 88.878%, 90.776%, 91.748%, 91.766%, 92.368%, and 92.777%. Similarly, for 90% of the training data, the recall attained by prevailing models is 87.876%, 91.876%, 92.876%, 93.008%, 93.544%, 94.368%, and 94.777%. Furthermore, using 90% of the training data, the existing models achieved F1 scores of 86.864%, 90.352%, 91.814%, 92.237%, 92.646%, 93.357%, and 93.766%. Likewise, the classical techniques reach a Cohen's Kappa for 90% of the training data of 0.739, 0.759, 0.777, 0.784, 0.795, 0.809, and 0.828. DHA-Net excels at focusing on the most relevant features, improving interpretability, and enhancing model precision. LSTM networks effectively capture temporal dependencies, making them ideal for sequential data like student interactions over time. By integrating DHA-Net's attention mechanism with LSTM's memory capabilities through Harmonic Analysis, the DHA-LSTM-FH Net leverages both spatial feature relevance and temporal dynamics. This combination results in more accurate and robust predictions of students' satisfaction, outperforming standalone models by effectively modelling complex patterns in educational datasets. From a practical perspective, the proposed framework can assist educators and platform developers in identifying patterns of student engagement within VR-based learning environments. The insights generated by the model can help instructors adapt teaching strategies, enhance content delivery, and design more personalized learning experiences. Furthermore, educational institutions can utilize the prediction outcomes to monitor student engagement levels in virtual classrooms and implement timely interventions to enhance overall learning effectiveness and student satisfaction.

Table 3. Comparative Discussion

Variations	Metrics	Hybrid CB-SEM and Deep ANN Model	PLS-SEM and ANN Models	Hybrid SEM-ML	ANN	TAM	E-SATNet	XGBoost	Proposed DHA-LSTM-FH Net
Dataset 1									
Training data of 90%	**Precision (%)**	85.876	88.878	90.776	91.478	91.766	92.368	92.777	***93.765***
	Recall (%)	87.876	91.876	92.876	93.008	93.544	94.368	94.777	***95.755***
	F1 Score (%)	86.864	90.352	91.814	92.237	92.646	93.357	93.766	***94.750***
	Cohen's Kappa	0.739	0.759	0.777	0.784	0.795	0.809	0.828	***0.838***
Dataset 2									
Training data of 90%	**Precision (%)**	86.876	87.877	88.765	89.257	89.765	90.478	91.777	92.654
	Recall (%)	86.790	88.877	91.876	92.368	92.665	93.467	93.889	94.765
	F1 Score (%)	86.833	88.374	90.294	90.786	91.192	91.948	92.821	93.698
	Cohen's Kappa	0.707	0.728	0.735	0.759	0.768	0.787	0.797	0.819

Since this research uses publicly available datasets, no direct human participants were involved in data collection. The datasets used in the experiments were obtained from open repositories and were already anonymized and prepared for research. Therefore, it did not involve any direct interaction with human subjects, and no personally identifiable information was accessed or processed. Nevertheless, the proposed framework is intended strictly for educational analytics and learning improvement. When deploying such systems in real educational environments, appropriate safeguards should be considered to protect student privacy and mitigate potential biases or over-reliance on automated predictions.

6. Conclusion

Predicting student satisfaction in VR teaching within the Metaverse uses data models to assess engagement and emotions. Existing models face challenges like diverse learning styles, tech accessibility, engagement consistency, platform usability, and data accuracy. To address these limitations, a robust model called DHA-LSTM-FH Net has been developed to predict student satisfaction in VR-based teaching within the Metaverse. The model utilizes data from Metaverse platforms, including virtual spaces, AR/VR devices, learning materials, and student information. Initially, interaction logs from VR sessions and student profiles are collected as input. The data is subjected to a sigmoid function to ensure consistency. Feature selection is conducted using Elastic Net and RFE to extract key attributes. LD-SMOTE is then applied to address data imbalance. Thereafter, student satisfaction is predicted using the DHA-LSTM-FH Net, which integrates DHA-Net and LSTM networks through Harmonic Analysis to enhance prediction accuracy. The DHA-LSTM-FH Net attained a precision of 93.765%, a recall of 95.755%, an F1 Score of 94.750%, and a Cohen's Kappa of 0.838. The proposed DHA-LSTM-FH Net demonstrates promising performance for predicting student satisfaction in VR-based learning environments; however, several limitations should be acknowledged. Attention mechanisms improve feature relevance, but the interpretability of DL models remains limited compared with simpler analytical approaches. Furthermore, the current framework primarily relies on interaction and behavioral data and does not incorporate emotional or physiological signals from VR environments. Future research can address these limitations by developing more computationally efficient architectures and integrating emotion detection technologies, such as VR headset signals, eye-tracking data, and facial expression analysis, to provide a more comprehensive understanding of student experiences.

Code availability statement: The source code for the proposed method is available at “https://github.com/shameen2718-png/DHA-LSTM-FH.git”.

References

[1] A. M. Al-Ansi, M. Jaboob, A. Garad and A. Al-Ansi, "Analyzing augmented reality (AR) and virtual reality (VR) recent development in education," Social Sciences & Humanities, vol. 8, no. 1, p. 100532, 2023.

[2] H. Rahman, S. A. Wahid, F. Ahmad and N. Ali, "Game-based learning in metaverse: Virtual chemistry classroom for chemical bonding for remote education," Education and Information Technologies, vol. 29, no. 15, pp. 19595-19619, 2024.

[3] S. Salloum, A. A. Marzouqi, K. Y. Alderbashi, F. Shwedeh, A. Aburayya, M. R. A. Saidat and R. S. Al-Maroof, "Sustainability model for the continuous intention to use metaverse technology in higher education: a case study from Oman," Sustainability, vol. 15, no. 6, p. 5257, 2023.

[4] S. M, A. R.A, K. J, M. I and A. M, "Modeling the Continuous Intention to Use the Metaverse as a Learning Platform: PLS-SEM and fsQCA Approach," In Current and Future Trends on Intelligent Technology Adoption, vol. 1, pp. 41-62, 2023.

[5] S. Mystakidis, "Metaverse," Encyclopedia, vol. 2, no. 1, pp. 486-497, 2022.

[6] Arpaci, I. and Bahari, M., "Investigating the role of psychological needs in predicting the educational sustainability of Metaverse using a deep learning-based hybrid SEM-ANN technique," Interactive Learning Environments, vol. 32, no. 6, pp. 2957-2969, 2024.

[7] Xiaojie Xu., and Yun Zhang., "Individual time series and composite forecasting of the Chinese stock index", Machine Learning with Applications, vol. 5, 2021.

[8] A. Marougkas, C. Troussas, A. Krouska and C. Sgouropoulou, "Virtual reality in education: a review of learning theories, approaches and methodologies for the last decade," Electronics, vol. 12, no. 13, p. 2832, 2023.

[9] Xiaojie Xu., and Yun Zhang., "An integrated vector error correction and directed acyclic graph method for investigating contemporaneous causalities", Decision Analytics Journal, vol. 7, 2023.

[10] A. Marougkas1, C. Troussas, A. Krouska and C. Sgouropoulou, "How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade," Multimedia Tools and Applications, vol. 83, no. 6, pp. 18185-18233, 2024.

[11] Xiaojie Xu., "Corn Cash Price Forecasting", American Journal of Agricultural Economics, vol. 102, pp. 1297-1320, 2020.

[12] C. Christou, "Virtual reality in education," In Affective, interactive and cognitive methods for e-learning design: creating an optimal education experience, pp. 228-243, 2010.

[13] M. Akour, H. A. Sghaier and O. A. Qasem, "The effectiveness of using deep learning algorithms in predicting students achievements," Indonesian Journal of Electrical Engineering and Computer Science, vol. 19, no. 1, pp. 387-393, 2020.

[14] Xiaojie Xu., and Yun Zhang., "Rent index forecasting through neural networks", Journal of Economic Studies, vol. 49, no. 8, pp. 1321 - 1339, 2021.

[15] Xiaojie Xu., and Yun Zhang., "Regional steel price index forecasts with neural networks: evidence from east, south, north, central south, northeast, southwest, and northwest China", The Journal of Supercomputing, vol. 79, no. 12, pp. 13601 - 13619, 2023.

[16] Mansi Aggarwal., and Vaibhav Vyas., "Data privacy preserved student career prediction with deep learning and blockchain based mechanism", Knowledge-Based Systems, vol. 337, 2026.

[17] V. Veeraiah, H. Khan, A. Kumar, S. Ahamad, A. Mahajan and A. Gupta, "Integration of PSO and deep learning for trend analysis of meta-verse," in 2nd International Conference Advanced Computing and Innovative Technologies in Engineering (ICACITE), 2022.

[18] Xiaojie Xu., and Yun Zhang., "Price forecasts of ten steel products using Gaussian process regressions", Engineering Applications of Artificial Intelligence, vol. 126, 2023.

[19] Xiaojie Xu., and Yun Zhang., "Corn cash price forecasting with neural networks", Computers and Electronics in Agriculture, vol. 184, 2021.

[20] Bingzi Jin., and Xiaojie Xu., "China commodity price index (CCPI) forecasting via the neural network", International Journal of Financial Engineering, vol. 12, no. 3, 2025.

[21] Xiaojie Xu., and Yun Zhang., "A high-frequency trading volume prediction model using neural networks", Decision Analytics Journal, vol. 7, 2023.

[22] T. Gaber, J. B. Awotunde, M. Torky, S. A. Ajagbe, M. H. and W. Li, "Metaverse-IDS: Deep learning-based intrusion detection system for Metaverse-IoT networks," Internet of Things, vol. 24, p. 100977, 2023.

[23] Zunaira Anwer., Shahnawaz Qureshi., Syed Muhammad Zeeshan Iqbal., Ali Zia., Sajid Anwer., "Predicting user behavior on video streaming by using watch-time duration analysis", Knowledge-Based Systems, vol. 332, 2025.

[24] T. Gaber, J. B. Awotunde, M. Torky, S. A. Ajagbe, M. H. and W. Li, "Metaverse-IDS: Deep learning-based intrusion detection system for Metaverse-IoT networks," Internet of Things, vol. 24, p. 100977, 2023.

[25] Xiaojie Xu., and Yun Zhang., "Machine Learning Coffee Price Predictions", Journal of Uncertain Systems, vol. 17, no. 4, 2024.

[26] Xiaojie Xu., and Yun Zhang., "Coking coal futures price index forecasting with the neural network", Mineral Economics, vol. 36, pp. 349 - 359, 2023.

[27] P. a. Rospigliosi, "Adopting the metaverse for learning environments means more use of deep learning artificial intelligence: this presents challenges and problems," Interactive Learning Environments, vol. 30, no. 9, pp. 1573-1576, 2022.

[28] X. Ding, F. Yang and F. Ma, "An efficient model selection for linear discriminant function-based recursive feature elimination," Journal of Biomedical Informatics, vol. 129, p. 104070, 2022.

[29] Ali, R.A., M. Soliman, M. R. Weahama, M. Assalihee and I. Mahmud, "Investigating continuous intention to use metaverse in higher education institutions: a dual-staged structural equation modeling-artificial neural network approach," Smart Learning Environments, vol. 12, no. 1, p. 3, 2025.

[30] ALMARZOUQI, A., A. ABURAYYA and S. A. SALLOUM, "Prediction of user's intention to use metaverse system in medical education: A hybrid SEM-ML learning approach," IEEE Access, vol. 10, pp. 43421-43434, 2022.

[31] Alshammari, S.H. and Alshammari, M.H., "Modelling the effects of perceived system quality and personal innovativeness on the intention to use metaverse: a structural equation modelling approach," PeerJ Computer Science, vol. 10, p. e2331, 2024.

[32] Akour, I.A., R. S. Al-Maroof, R. Alfaisal and S. A. Salloum, "A conceptual framework for determining metaverse adoption in higher institutions of gulf area: An empirical study using hybrid SEM-ANN approach," Computers and Education: Artificial Intelligence, vol. 3, p. 100052, 2022.

[33] Al-Adwan, A.S., N. Li, A. Al-Adwan, G. A. Abbasi, N. A. Albelbisi and A. Habibi, "Extending the technology acceptance model (TAM) to Predict University Students' intentions to use metaverse-based learning platforms," Education and Information Technologies, vol. 2, no. 1, pp. 486-497, 2022.

[34] Manikandan Suriyanarayanan., Aranganathan.P., Nagaraj Navalgund., Preethi Baligar., Sanjay V Hanji., and Shashidhar S. Mahantshetti., "Unveiling Student Satisfaction in Online Learning: Leveraging Artificial Neural Networks for Predictive Insights", Journal of Engineering Education Transformations, vol. 38, 2025.

[35] Sulis Sandiwarno., Dana Indra Sensuse., Harry Budi Santoso., Deden Sumirat Hidayat., Ally S. Nyamawe., and Abdallah Yousif., "E-SATNet: Evaluating Student Satisfaction with Lecturer Responses in Asynchronous Online Discussions Using Sentiment and Semantic Similarity Analysis", Big Data and Cognitivie Computing, vol. 9, no. 228, 2025.

[36] Nisar Ahmed Dahri., Fida Hussain Dahri., Asif Ali Laghari., and Muhammad Javed., "Decoding ChatGPT's impact on student satisfaction and performance: a multimodal machine learning and explainable AI approach", Complex Engineering Systems, vol. 5, no. 7, 2025.

[37] Durgesh Kumari., and Gyanendra Singh., "Study of Heart Disease Prediction Using CNN Algorithm", Journal of Emerging Technologies and Innovative Research (JETIR), vol. 8, no. 7, 2021.

[38] F. Amini and G. Hu, " A two-layer feature selection method using genetic algorithm and elastic net," Expert Systems with Applications, vol. 166, p. 114072, 2021.

[39] J. Lyu, J. Yang, Z. Su and Z. Zhu, "LD-SMOTE: A Novel Local Density Estimation-Based Oversampling Method for Imbalanced Datasets," Symmetry, vol. 17, no. 2, p. 160, 2025.

[40] M. Waqas, A. Ahmed, T. Maul and I. Y. Liao, "Enhancing breast cancer histopathological image classification using attention-based high order covariance pooling," Neural Computing and Applications, vol. 36, no. 36, pp. 23275-23293, 2024.

[41] S. S. Namini, N. Tavakoli and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in International Conference on Big Data (Big Data), 2019.

[42] E. D. and E. S., "Estimation of trigonometric components in time series," Journal of the American Statistical Association, vol. 77, no. 378, pp. 381-387, 1982.

[43] "The impact of Virtual Reality on the Education dataset," [Online]. Available: https://www.kaggle.com/datasets/waqi786/impact-of-virtual-reality-on-education [Accessed February 2025].

[44] "VR_College_Teaching_Dataset," [Online]. Available: https://www.kaggle.com/datasets/ziya07/vr-college-teaching-dataset [Accessed February 2025].

[45] P. Misra and A. S. Yadav, "Improving the classification accuracy using recursive feature elimination with cross-validation," International Journal on Emerging Technologies, vol. 11, no. 3, pp. 659-665, 2020.

[46] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," Neurocomputing, vol. 323, pp. 203-213, 2019.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.