International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

An Efficient Student Mental Health Analysis Method Based on Recurrent Neural Networks

S. Anitha

Department of Computer Science and Applications, Vivekanandha College of Arts and Sciences for Women (Autonomous), Elayampalayam, Tamil Nadu, India
selvianitha1974@gmail.com

DOI: https://doi.org/10.47760/ijcsmc.2026.v15i07.005

Abstract: Today's college students will undoubtedly experience stress. When under stress, a person may exhibit potent emotional and behavioral reactions. Stress-related mental health problems are among the most common causes of stress for college students worldwide. Research on the effects of particular activities, such as study trips, on students' mental health and how these activities might be tracked using cutting-edge technologies is lacking, though. Deep learning has lately found broad use in college students' mental health (SMH) education and management due to their capacity to evaluate, categorize, and notify psychological data with high quality. The present research work focused on three recurrent neural network (RNN) methods that were used for the analysis of student mental health issues. This analysis uses three types of RNN methods such as long short-term memory (LSTM), gated recurrent neural network (GRU), and simple RNN (SRNN) for analyzing student mental health. The dataset is collected from various engineering college students to analyze their SMH issues. Examining the experimental outcomes offered on test data based on f-score, recall, accuracy, and precision.

Keywords: Student mental health analysis; Student performance analysis; Long-short term memory; Gated recurrent neural network; Recurrent neural network; social media

I. Introduction

Social media provides chances for communication, connection, and even learning, but if it is not properly controlled, it can have a very negative impact on students' academic achievement. To make social media a tool that enhances learning rather than a barrier to academic success, students must learn how to balance their online activities with their academic obligations. Parents and teachers are crucial in helping adolescents responsibly use social media. Social media can affect kids' academic achievement in both positive and negative ways [1, 2]. It presents several distractions that can impair a student's performance and focus, even while it can also present chances for learning, networking, and involvement. This is a summary of how social media can impact academic achievement. Because Facebook, Instagram, TikTok, and Twitter are meant to be entertaining,

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

students frequently spend too much time looking through their feeds, watching videos, or conversing with friends on these sites [3]. This may lead to procrastination and a postponement of finishing work or studying for tests. Study periods can be disrupted by social media updates, which impair focus and concentration. Students may find it challenging to focus fully on their initial assignments after even a little diversion, distraction, mental health issues, and misinformation—need to be carefully managed. Balancing social media usage with other aspects of life is essential to ensure its positive impact [4] . Figure 1 shows the framework for student performance prediction based on social media.

Deep learning (DL) has significant potential for analyzing and improving student mental health. By leveraging advanced algorithms and vast amounts of data, deep learning techniques can help identify early warning signs, predict mental health issues, and provide tailored interventions for students [5, 6]. Recurrent neural networks (RNNs) are quite useful for comprehending and forecasting student mental health because they are especially well-suited for evaluating time-series or sequential data. By taking into account the temporal relationships between various data sets, RNNs can monitor the progressive or cyclical changes in mental health over time. The benefit of employing RNN-based classifiers is that these DL networks may briefly remember past states and use them in the current calculation [7]. Hence, the present research work uses three different RNN methods such as SRNN, LSTM, and GRU for analyzing the student's mental health and how it will affect their academic performance. The contributions of the research work are as follows,

- The RNN method is developed for analyzing student performance based on students' mental health based on their social media activities
- Collecting the student data from various engineering colleges in Rajasthan, India
- Different preprocessing methods are considered to enhance the prediction models
- Evaluating the performance of models using various performance measures such as accuracy, precision, recall, f-score, and computation time.
- Three kinds of RNN methods are used such as SRNN, LSTM, and GRU based on their different weight ranges and different hidden layer size.
- The optimal weights and hidden layers are considered to test the data to analyze the performance.
- The numerical datasets are used to analyze the student's mental health based on social media behavior.

The paper is organized as follows: Section 2 discusses the purpose of the research methods; Section 3 discusses related work; Section 4 discusses research methodologies; Section 5 discusses the analysis of experimental results; and Section 6 draws the paper's conclusion.

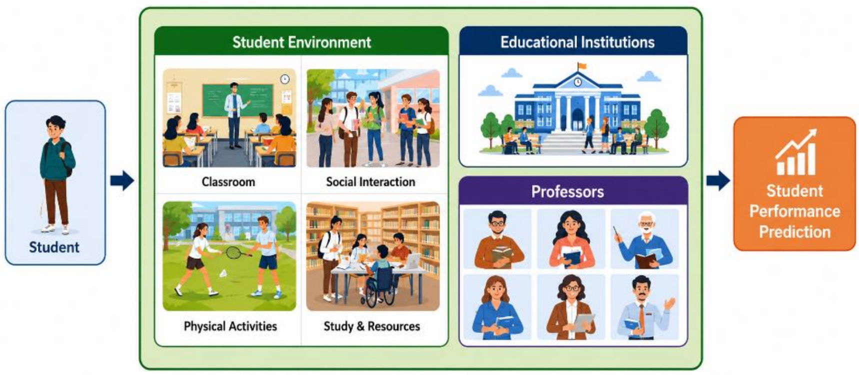

Fig. 1 : Framework for Student Mental Health Analysis

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

II. Related works

The advancement of research, the improvement of techniques, and the efficacy of interventions all depend on related work in student mental health analysis. Researchers, educators, and mental health practitioners can learn more about the difficulties kids encounter, the efficacy of various strategies, and how technology and data can be used to enhance mental health outcomes by examining and expanding on previous research. In 2026, Bharathi et al. [8] proposed a hybrid Weighted Extreme Learning Machine-Coati Optimization Algorithm (WELM-COA) model for prediction of student performance. The study optimized the input weights and hidden layer biases of ELM and WELM using COA to improve the prediction accuracy and to solve the problem of random initialization. The experimental results on benchmark student datasets indicated that the proposed WELM-COA model performed better than Spike Neural Network (SNN). The method, however, mainly focuses on optimizing the parameters and does not capture the complex temporal learning patterns or the sequential behavior of the students well. J. Li et al. [9] (2026) proposed an interpretable belief rule base model (IBRB-m) for student performance prediction. The framework used Random Forest for feature selection and incorporated interpretable constraints into multi-parameter optimization to enhance model transparency. The proposed approach achieved reliable prediction performance with maintaining interpretability. But the model is rule-based and might face challenges in modeling highly nonlinear relationships in large educational datasets.

Buzducea et al. (2026) [10] investigated students' perceptions on machine learning technologies for planning, monitoring and prediction of academic performance. Descriptive statistics, inferential analysis, regression models, exploratory factor analysis and clustering were used in the study, which also discussed the advantages and ethical issues of ML adoption in education. The work offers valuable insights into student attitudes but it does not present or evaluate a predictive machine learning framework for academic performance. Wang et al. (2025) [11] presented a machine learning framework to predict online student performance from behavioral learning indicators. We applied correlation analysis to identify significant learning features and used a Taylor expansion to build an improved logistic regression model. The experimental results indicated better prediction performance than conventional models and emphasized the importance of learning initiative and study duration. However, logistic regression's linearity limits its capacity to capture intricate nonlinear relationships among educational features. Arun et al. (2025) [12] developed an RBF-ELMAN neural network model to examine the impact of students' social media use on academic performance. The framework contained preprocessing, feature extraction and feature selection before training the hybrid neural network. The framework was able to achieve an accuracy of 96.35%, which was better than the conventional RBF and ELMAN models. The proposed approach shows promising performance, but it relies on handcrafted feature engineering and may not generalize well across different educational datasets with different characteristics.

Venkatachalam et al. (2023) [3] suggest using the Distinctive Deep Learning (SADDL) model to predict student accomplishment. Initially, the student's academic and demographic data are gathered, as are the posts they make on online social networks regarding their academic achievement. The physiological characteristics are extracted from the data of internet posts using a Latent Dirichlet Allocation (LDA). After that, the SADDL network—which consists of three modules—is given all the attributes of the student's data. The LSTM module is used to learn the temporal dependency, the multidimensional Deep CNN (DCNN) module is used to learn the correlation attributes, and the Multi-Layer Perceptron (MLP) module is used to predict the academic performance of the students. Lastly, the experimental findings demonstrate that the SADDL has a 91.71% accuracy rate in predicting students' academic success. K. Li et al. (2024)[13] suggest the College Student Mental Health Prediction Model (BDNN-CSMHPM), which is built on biosensors and DNNs, to identify college students' mental stress levels while they are on study trips. This model uses the best BDNN to classify the mental health condition as normal, negative, or positive based on biosensor data, such as EEG and biomechanical parameters. The collected emotional and biomechanical data is then classified using BDNN, and the emotional state of college students is ascertained from the classification results. To accurately capture the

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

information in the original EEG data, many biosensor characteristics must be extracted, as distinct features may reflect different elements in the original data. Second, the auto-learn model integration approach relies on combining different features. Third, the combined features are put into the BDNN, which classifies emotions.

Y. Ding, X. Chen(2020) [14] suggest using text-level mining of college students' Sina Weibo data to identify sadness in this population. First, gather text data from Sina Weibo users who are college students, then turn that data into input data for machine learning. For feature extraction, DNNs are employed. To classify the input data and ultimately achieve the recognition of sadness, the deeply integrated support vector machine (DISVM) technique is presented. To a certain degree, DISVM increases the stability of the recognition model and boosts the diagnosis accuracy of depression. Through the use of Sina Weibo data, simulation tests confirm that the suggested depression detection technique is capable of identifying possible depression patients among college students. A. S. Aljaloud *et al*. (2022) [15] analyzed four DL models as part of a mixed-methods study to forecast student performance. Reports on seven general preparation courses provided the data. To determine potential predictive KPIs related to the electronic Blackboard report, they were examined utilizing a documentary analysis methodology. The degree to which these elements are linearly connected with students' performance indicators was investigated using correlational analysis. The results showed that, out of the four models examined, a predictive model that integrated convolutional neural networks and LSTM (CNN-LSTM) was the best approach. The primary inference made from this data is that the CNN-LSTM combination may result in interventions that maximize and increase university utilization of Blackboard learning.

V. Das Swain et al. (2024) [16] Examine the feasibility of employing language on private social media as an indicator of crisis-related disruptions in mental health. Over two years, 4124 Facebook posts made by 43 undergraduate students were modeled. Over time, they noticed trends in the psycholinguistic elements of their posts and comments. These trends demonstrated the impact of COVID-19 on their mental health. The improvement over a model that was trained using the participant's statements about their mental health. Data minimization became possible when it was found that selecting the periods seven months after the COVID-19-induced lockdown yielded superior findings. Other mental states, such as anxiety, depression, social isolation, and suicidal conduct, could also be predicted by the features used. Developed a machine learning model to predict COVID-19-induced mental health problems using language from private social media. I. J. Ratul et al. (2023) [17] use supervised machine learning techniques to develop a machine learning-based approach for analyzing student well-being. The chi-squared test and Principal Component Analysis (PCA) were used as feature reduction methods. Additionally, for hyperparameter optimization (HPO), Grid Search Cross-Validation (GSCV) and Genetic Algorithm (GA) were used. When used in conjunction with GSCV for HPO and PCA as a feature reduction technique, the Multilayer Perceptron model demonstrated the highest accuracy. This study's convenience sampling method solely takes self-reported data into account, which could lead to biased and non-generalizable conclusions. Future studies should examine a sizable data set and concentrate on monitoring the long-term effects of therapies and coping mechanisms. The study's findings can be applied to the creation of plans to lessen the negative consequences of excessive mobile device use and to enhance students' well-being in times of stress, such as pandemics. Q. Chen et al. (2023) [18] created a set of tests that were carried out to assess students' stress levels by having them play Sudoku games in various distracting environments. To evaluate stress levels, the physiological signals that were gathered—such as PPG, ECG, and EEG—were analyzed using improved models like LRCN and self-supervised CNN. Following the studies, the results were contrasted with the subjects' self-reported stress levels. The results show that the improved models introduced in this study are highly effective at determining stress levels.

S. Pandey et al. (2023) [19] created two chatbots with six designs each, one generative and one retrieval-based. The accuracy of the retrieval-based chatbots is 83.22% for Vanilla RNN, 89.87% for LSTM, 91.57% for Bidirectional LSTM (Bi-LSTM), 65.57% for GRU, and 82.33% for CNN. By contrast, the encoder-decoder designs of generative-based chatbots are 94.45% accurate. The most important difference is that retrieval-based chatbots can only reply to inputs that best match the outputs they already know, but generative-based chatbots

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

can create new content. N. Kirlic *et al. (2023)* [20] investigated correlates of SBQ-R scores and built a machine learning (ML) pipeline using stacking and tiered cross-validation. The SBQ-R found that 9.6% of pupils were at significant risk for STBs. The ML algorithm was able to explain 28.3% (95%CI: 28–28.5%) of the variance in baseline SBQ-R scores, with the most significant components being positive affect, social isolation, depression severity, and meaning and purpose in life. Only 1.8% of students were found to be at high risk at the end of the year, indicating a considerable decrease in STBs. In addition to identifying new, potentially changeable characteristics including positive affect and social connectivity, analyses confirmed established parameters linked to STBs during the first semester of college. L. Chen et al. (2024) [21] developed a new DeepPsy-based method for detecting students with mental health problems was suggested following the improvement of the algorithm proposed in the previous phase. To construct the Internet trajectory matrix and recover the daily Internet trajectory pattern, a two-dimensional convolutional neural network was employed. Each day's time-dependent connection was recorded using a long-term memory network and the Internet trajectory pattern was fused with basic attributes to construct a deep learning network. The DeepPsy methodology identified 75.5% of students with mental health problems through experimental research. Teachers can then utilize the methodology proposed in this article to identify students who might be experiencing untreated mental health problems.

Q. Gao et al. (2024) [22] study the results of college students who take required or elective courses on mental health education. Last but not least, research on the mental health of Chinese college students is conducted in light of the contemporary circumstances in China. This includes statistics, analysis, and research on Chinese college undergraduates. College students can take required and elective courses on mental health education at the schools and colleges that have been evaluated. The student questionnaire survey indicates that 86.7% of participants think it is essential to create educational courses related to mental health, 61.9% think that such courses should be made mandatory, and students want to incorporate group projects or activities into the teaching process to enhance their engagement and experience. J. Luo et al. (2023) [23] produces the DNN-MHE model to assess the impact of mental health education. Data on the knowledge, attitudes, and behaviors of 916 students regarding mental health were gathered by a questionnaire survey. Five completely connected layers are used by DNN-MHE to forecast mental health measures. Experiments show that DNN-MHE outperforms RNN, CNN, and shallow MLP models with an accuracy of 99.46%. The effects of training iterations, neuron count, and data sample count on performance are validated by ablation studies. All things considered, DNN-MHE makes it possible to analyze mental health education accurately and effectively, which has useful ramifications for enhancing academic programs.

III. RESEARCH METHODS

Recurrent neural networks (RNNs) are not limited to processing input in a single direction, unlike simple neural networks like Multilayer Perceptrons (MLPs)[24]. In addition to looping across layers, RNNs can temporarily store information for use at a later time. A conventional neural network is shown by the symbol NN, where the input is represented by x_p and the output by h_p. Since the input is processed across multiple layers, RNNs are by definition regarded as deep neural networks. To demonstrate the depth of the RNN design, a conventional SRNN is unrolled, as seen in Figure 2. Despite their effectiveness in a variety of prediction tasks, SRNNs suffer from a vanishing gradient issue. Other RNN designs, including the GRU and LSTM, were developed to address this problem. Let x_p represent the input to the SRNN approach. The weight of the connection between the input x_p and the hidden layer is denoted by w_{hx}. The weight of the link relating the current hidden layer with the following hidden layer is represented by w_{hh}. The weight between the last hidden layer with its corresponding output layer is denoted by w_{hy}. Let h_p symbolize the output at time p. The biases added to the connections between the hidden and the output layer are denoted by b_h and $b_{y,}$ respectively. An SRNN is statistically created as follows:

$$h_p = g(h_{p-1,} x_p) \tag{1}$$

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

$$h_p = g([w_{hx} \cdot x_p + w_{hh} \cdot x_{p-1,}] + b_h) \tag{2}$$

$$\bar{y}_p = g([w_{hx} \cdot x_p + w_{hh} \cdot x_{p-1,}] + b_h) \tag{3}$$

Where $\bar{y}_p$ is the predicted result. At each step p , the error (E) between the prediction and the actual output values is calculated. In ML terminology, L is called Loss, L. The ultimate aim of the training process is to decrease L. In other words, a respectable model will possess a minimal L. For a given dataset, the L, produced by training of one batch of data at time, p , is considered in the following expression:

$$L(\bar{y}_p, y) = \sum_{p=1}^{n} L(\bar{y}_p - y_p) \tag{4}$$

Where N represents the maximum training time. SRNN has a reputation for suffering from a vanishing gradient issue or an explosion of the gradient during the training procedure. These problems always occur as the size of a given dataset expands and as N increases. The vanishing and exploding gradient problem are addressed by the introduction of LSTMs [25] and GRUs [26]. Figures 3 and 4 show the architecture of LSTM and GRU respectively. The expressions provide a mathematical formulation of the flow of information within a LSTM and GRU, respectively.

$$\text{LSTM Formulation=} \begin{cases} h_p = \sigma(W_{g.} [v_{p-1}, \; x_p] + b_g \\ k_p = \sigma(W_{k.} [v_{p-1}, \; x_p] + b_p \\ S_{\acute{p}} = \tan h \, (W_{s.} [v_{p-1}, \; x_p] + b_s \\ S_p = g_p * S_{p-1,} + p_t \; * S_{\acute{p}} \\ r_p = \sigma(W_{r.} [v_{p-1}, \; x_p] + b_r \\ v_p = r_p * \tan h \, (S_p), \end{cases} \tag{5}$$

Where S denotes the cell state. The following activation functions are used in the LSTM unit: hyperbolic tangent, $\tan h$, (a)=$\tan h(a) = \frac{1-e^{-2a}}{1+e^{-2a}}$, and the sigmoid, $\sigma(a) = \frac{a}{1+e^{-a}}$. x represents the input vector. v *denotes* the output vector. W and p represent their related biases.

$$\text{GRU Formulation:} \begin{cases} k_p = \sigma(W_{k.} [S_{p-1}, \; x_p] + x_p \\ z_p = \sigma(W_{z.} [S_{p-1}, \; x_p]) \\ S_{\acute{p}} = \tan h \, (W_{.} [\, z_p * S_{p-1}, x_p \,]) \\ S_p = (1 - z_p) * S_{p-1,} + k_p * S_{\acute{p}} \end{cases} \tag{6}$$

Where x is the input vector. S_p is the computed prediction. k_p denotes the update function. The corresponding weight is defined by w as in the LSTM, the GRU uses the hyperbolic tangent as well as the sigmoid activation function to squash the information.

IV. EXPERIMENTAL ANALYSIS

This paper's primary objective is to create predictive models that use deep neural networks and some conventional machine-learning approaches to forecasting students' academic performance in future courses early in the semester based on their prior academic performance. Figure 5 is the flow diagram of the proposed research work. Students will benefit from these models since they will be notified early in the semester of their likely outcomes. As a result, individuals can improve their academic performance after the term. MATLAB

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

2019 is employed to carry out the experimental process. All simulation is implemented on an HP laptop on Windows 11 with i5-1235U, 1.30 GHz, and 8 GB RAM.

4.1 Data collections

The datasets were collected from various engineering colleges in Rajasthan, India. We are making the questionnaire pattern in Google form which is shared with undergraduate (UG) and postgraduate (PG) engineering college students. A total of 1828 data samples were collected which are used for analysis purposes. Details descriptions of the data samples are discussed in Table 1. Totally 15 features were collected, and 14 features were considered as inputs to the classification such as Age, Gender, Education level, Usage of social media, frequently used platform, Average time (usage of social media per day), Headache, Depression, burning eyes, Sleeping, Behavioural disorder, Anxiety, Dullness, studying (hours per day) for both binary and multiclassification model and 15th features is considered as target values such as Last Semester GPA.

Table 1. Descriptions of datasets

S.No	Features	Descriptions
	Age	Numeric
	Gender	Male =0; Female =1
	Education level	UG =1; PG =2
	Usage of social media	Facebook =1; Twitter (X) = 2; YouTube = 3 Instagram =4; WhatsApp=5; Other=5
	Frequently used platform	String
	Average time (usage of social media per day)	Time
	Headache	Yes=1; No =0
	Depression	Yes=1; No =0
	Burning eyes	Yes=1; No =0
	Sleeping	Yes=1; No =0
	Behavioural disorder	Yes=1; No =0
	Anxiety	Yes=1; No =0
	Dullness	Yes=1; No =0
	Studying (hours per day)	Time (Hours)
	Last Semester GPA:	Final Semester Mark (CGPA) Fail = 0; Pass=1

4.1 Data preprocessing

Different preprocessing steps such as data cleaning, Discretization, feature scaling, and validation methods are considered in the work for enhancing the quality of datasets and performance of prediction models. The following subsections are discussed as,

Table 2. Process of discretization

Student marks	Binary class label	Multi-class label
91–100%	Pass	Outstanding
81–90%		Excellent
71–80%		Very good
61–70%		Good
51–60%		Above average

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)

ISSN: 2320-088X
Impact Factor: 7.056

51–60%		Average
40–50%		Pass
< 40	Fail	Fail

a) *Data cleaning*
Our actual data contains some missing numbers, such as the absence of a student from any exam. As a result, we eliminated the missing data using this procedure.

b) *Data Discretization*
It is one of the most crucial phases in machine learning since it transforms the unprocessed data into an appropriate format to fix flaws in the real-world dataset and produce improved outcomes[27]. We took the subsequent actions: the step of data pre-processing, which covers handling imbalanced datasets, feature encoding, data cleansing, and data discretization, to describe our dataset as a classification issue, we employed the discretization technique in this study to translate the numerical values of the students' grades into nominal values. We employed the equal-width binning technique to carry out this phase by the university's standard grading, as indicated in Table 2. To make it easier for the algorithm to learn and produce better results, we separated the input features into five nominal intervals based on the student's grades: Outstanding, Excellent, very good, good, above average, Average, Pass, and Fail. Furthermore, we separated the class label into two nominal intervals, Pass and Fail, according to the grades of the pupils. The discretization step results are shown in Table 3. Using the frequent distribution of categorical variables, we employed the chi-square test as a statistical tool to determine the association between them. The chi-square test postulates that category variables do not correlate with one another. The significance level for this test was set at 0.05. A statistical association between categorical variables was present if the hypothesis was disproved. The correlation between category variables was determined using the chi-square as explained below.

$$x^2 = \frac{(o-e)^2}{2} \quad (7)$$

where the predicted frequency is denoted by e and the observed frequency by e. Using the chi-square test, we discovered a significant relationship between the student's academic performance in the data structure course, which is the next course of the second year, and their grades in each of the first year's courses.

c) *Feature encoding*
Our features are currently categorical data, and categorical data is incompatible with machine learning methods. As a result, before entering the input characteristics into the model, we must transform this categorical data into numerical form. In this work, we transformed our categorical data into a numerical format using an encoding method known as label encoding. Using this method, each nominal variable is given an integer value. Integer numbers are generated from our input features. We represent the numerical values for multi-class labels as follows: Outstanding (0), Excellent (1), Very Good (2), Good (3), Above Average (4), Average, (5) Pass (6), and Fail (7). Furthermore, the output feature was transformed into whole numbers. Fail is one, and pass is zero.

d) *Handling imbalanced dataset*
Our dataset was extremely unbalanced after the discretization phase was applied, and the distribution of students' class labels according to their grades was not equal. Class ``Pass'' contains more samples in our class label than class ``fail'', which has fewer samples. Figure 6 shows the percentage of data samples based on classes. The class label distributions (pass (1509), fail (319)) were 83 % percent and 17 % percent, respectively. This is a serious difficulty that arises in classification issues and results in subpar performance. One of the most crucial elements in enhancing the models' performance is resolving the imbalanced dataset issue. The majority class dominates the minority class as a result of this issue. As a result, the classifiers typically belong to the majority class, and their performance is untrustworthy. We must thus find a solution to this issue since it could produce false results. To address the issue of an unbalanced dataset, the oversampling method creates extra

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

instances to increase the number of minority classes in the dataset. This works well with tiny amounts of data. The synthetic minority oversampling technique (SMOTE) [28] is used to creates fresh minority class samples to balance out the class frequencies, which helps to alleviate this problem.

e) Feature scaling

Scaling the dataset's input features within a narrow range is crucial. This stage is crucial since it speeds up the algorithm's learning process [29]. We employed a conventional scaler, which is one of the most widely used feature scaling algorithms. One of the reasons our method achieves improved accuracy is that we applied this strategy to every input feature in the dataset. We determined the mean and standard deviation for each feature to standardize it in the input variables. Next, each sample X's new value of X_s is determined as follows:

$$X_s = \frac{x-\mu}{\sigma} \tag{8}$$

where x and μ stand for a feature's mean and standard deviation, respectively.

f) Validation

Since the complete dataset is used for both training and testing, it is an upgraded type of cross-validation technique used to improve the performance of machine learning techniques. When the dataset is really little, as it was in our study, it is also a helpful tool. Using this method, the dataset is divided into K equal-sized sections at random. To construct the model, one subset was utilized for testing, and the remaining subsets were used for training. This procedure is occasionally repeated, and the model's final output is derived from the testing set's average result[30]. We employed stratified K-fold, a kind of K-fold cross-validation, in our investigation. Because the class distribution in each fold is created using the same number of samples for each class, it is employed with classification problems. A stratified 10-fold cross-validation was employed. Then, 8-fold data samples are used for training and 2-fold are used for testing.

4.2 Parameters setting

The size of the hidden layers in each algorithm is three. $Sigmoid$ and $ReLU$ are used as the activation functions. Both the binary and multiclass classification tasks were used in the initial round of the trials. Three different RNN types—SRNN, LSTM, and GRU—were implemented. The hyperparameters were as follows:

- Binary classification: $loss$ = 'binary_crossentropy', $optimiser$ = 'sgd' (Stochastic Gradient Descent), initial $learning\_rate$ = 0.05. The dense layer is 3.
- Multi-classification: $loss$ = 'categorical_crossentropy', $optimiser$ = 'adam', initial $learning\_rate$ = 0.051. The dense layer is 3.

4.3 Performance metrics

Several measures are used to assess the models' performance while analysing student mental health using deep learning. Four performance metrics are considered in this research work such as accuracy, precision, recall, and F1-Score are computed using a confusion matrix were used to assess the performance of the prediction algorithm [31].

Accuracy is calculating the proportion of examples that are correctly classified which is defined as follows,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{9}$$

Precision shows the percentage of accurate identifications which is defined as follows,

$$Precision = \frac{TP}{TP+FP} \times 100 \tag{10}$$

Recall is evaluating the capacity to find all pertinent examples.

$$Recall = \frac{TP}{(TP+FN)} \times 100 \tag{11}$$

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

F-measures is the harmonic mean of precision and recall which is defined as follows,

$$F1 - Score = \frac{(2 \times Precision \times Recall)}{(Precision \times Recall)} \tag{12}$$

Where, TP is number of students correctly classified as pass, FP is number of students incorrectly classified as pass when actually they fail, TN is number of students correctly classified as fail and FN is number of students incorrectly classified as fail when actually they pass.

4.4 Results analysis and discussions

Validating the findings, comprehending their practical consequences, and pinpointing opportunities for further research all depend on the examination and discussion of experimental results. When it comes to student mental health analysis, they help close the gap between unfiltered performance stats and useful insights. RNN weights and hidden neurons are essential for assessing students' mental health, particularly when working with sequential data. The neural network's trainable parameters that control how strongly neurons link to one another are called weights. Weights specify how input data is converted into outputs by the network. The network modifies weights during training to extract pertinent information from the data. For instance, weights could draw attention to characteristics like a propensity for using derogatory language or erratic sleeping habits in mental health analyses. The layers in between the input and output layers are known as hidden neurons or units. They carry out calculations that convert incoming data into a format that the following layer may use. Raw data is gradually transformed into higher-level, more abstract qualities by hidden layers. Deeper layers in student mental health may be able to identify complicated conditions like chronic stress, while early levels may be able to identify fundamental patterns like inconsistent sleep. Weights and hidden neurons collaborate in RNN to change input data, extract features, and generate predictions. Through training weight adjustments, the network gains the ability to generalize from examples and recognize intricate patterns in a variety of datasets. The present research work focussed on different ranges of weights (between 0 and 1, -0.5 and +0.5, and -1 and +1), different hidden neurons (such as 20, 40, 60, 80, and 100) in each hidden layer of RNN methods such as SRNN, LSTM, and GRU. Finally, optimal weights and hidden neurons are considered for student health analysis on test data. Classification models are essential for classifying students according to their mental health state in student mental health analysis. The complexity of the task and the intended results determine whether to use binary or multiclass categorization.

a) Binary classification

Binary classification is the process of classifying pupils into two groups according to a particular standard, like if they achieve a certain level of performance. When making high-level decisions, such as determining which kids are at risk of failing, binary categorization works well. Tables 3, 4, and 5 and Figures 2,4 and 6 show the performance results based on binary classification, and Figures 3,5 and 7 show the computational weights based on different weight ranges and hidden neurons in the hidden layers for LSTM, SRNN, and GRU, respectively. According to the binary classification, the LSTM produced optimal results such as an accuracy of 93.86 % when choosing 40 hidden neurons with weights ranging between 0 and 1, an accuracy of 92.79 % when choosing 60 hidden neurons with weights between -1 and 1, and 92.07 % when choosing 40 hidden neurons weights between -0.5 and 0.5. However, SRNN is taking a low computational time compared to other methods such as LSTM and GRU. According above results, LSTM produced optimal results with 40 hidden neurons with weights ranging between 0 and 1.

Table 3. Performance results of binary classification – weights between 0 and 1

Methods	HS	Accuracy	Precision	Recall	F1 –score	Time (Sec)
LSTM	20	91.58	91.22	91.91	91.56	9.46
	40	**93.86**	**93.31**	**94.4**	**93.85**	8.62
	60	92.81	92.09	93.51	92.79	9.14

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

	80	93.61	92.79	94.43	93.6	9.02
	100	93.65	93.26	94.04	93.65	9.32
GRU	20	89.79	89.29	90.23	89.76	14.2
	40	91.11	90.99	91.21	91.1	13.45
	60	**91.25**	**90.73**	**91.74**	**91.23**	14.55
	80	91.23	90.73	91.69	91.21	13.88
	100	90.51	90.24	90.76	90.5	15.1
SRNN	20	85.75	84.66	86.66	85.65	4.5
	40	86.11	85.28	86.85	86.03	4.56
	60	86.43	85.6	87.12	86.36	3,25
	80	**88.68**	**87.73**	**89.52**	**88.61**	4.88
	100	88.32	87.28	89.23	88.25	4.54

Table 4. Performance results of binary classification – weights between -1 and 1

Methods	HS	Accuracy	Precision	Recall	F1 –score	Time (Sec)
LSTM	20	90.05	89.39	90.66	90.02	11.07
	40	90.65	90.13	91.14	90.63	10.92
	60	**92.79**	**92.06**	**93.51**	**92.78**	9.03
	80	90.47	90.02	90.88	90.45	9.55
	100	92.01	91.02	92.98	91.98	11.42
GRU	20	88.13	87.86	88.36	88.11	13.24
	40	89.17	88.12	90.13	89.11	12.55
	60	89.78	88.88	90.61	89.73	13.46
	80	**90.76**	**90.17**	**91.32**	**90.74**	12.73
	100	89.52	88.84	90.16	89.49	13.08
SRNN	20	86.41	85.58	87.13	86.34	5.14
	40	86.84	86.17	87.43	86.79	4.26
	60	86.23	84.91	87.37	86.12	3.68
	80	87.66	87.41	87.89	87.65	4.96
	100	**87.86**	**87.57**	**88.12**	**87.84**	4.58

Table 5. Performance results of binary classification – weights between -0.5 and 0.5

Methods	HS	Accuracy	Precision	Recall	F1 –score	Time (Sec)
LSTM	20	91.51	90.8	92.19	91.49	10.07
	40	**92.07**	**91.8**	**92.32**	**92.06**	9.43
	60	91.73	91.12	92.3	91.71	8.66
	80	91.33	90.91	91.72	91.31	9.25
	100	90.93	90.12	91.7	90.9	9.44
GRU	20	88.06	87.42	88.63	88.02	14.26
	40	89.74	88.83	90.59	89.7	13.58
	60	**90.74**	**90.33**	**91.13**	**90.73**	12.47
	80	89.35	88.38	90.24	89.3	12.05
	100	88.38	87.75	88.94	88.34	13.45
SRNN	20	85.11	84.61	85.52	85.06	6.75
	40	86.24	85.71	86.68	86.19	5.84
	60	86.68	86.17	87.12	86.64	6.1
	80	**88.11**	**87.37**	**88.77**	**88.06**	6.47
	100	87.68	86.91	88.36	87.63	6.41

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84

(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)

ISSN: 2320-088X
Impact Factor: 7.056

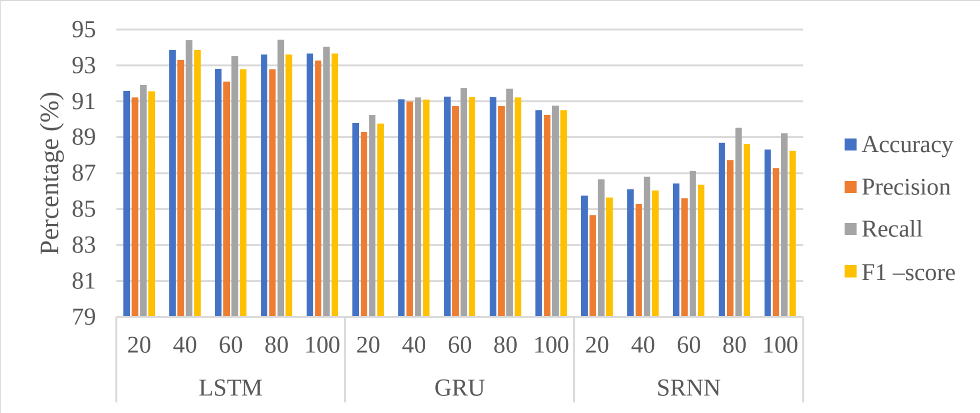

Fig. 2 : Performance results of binary classification - weights range between 0 and 1

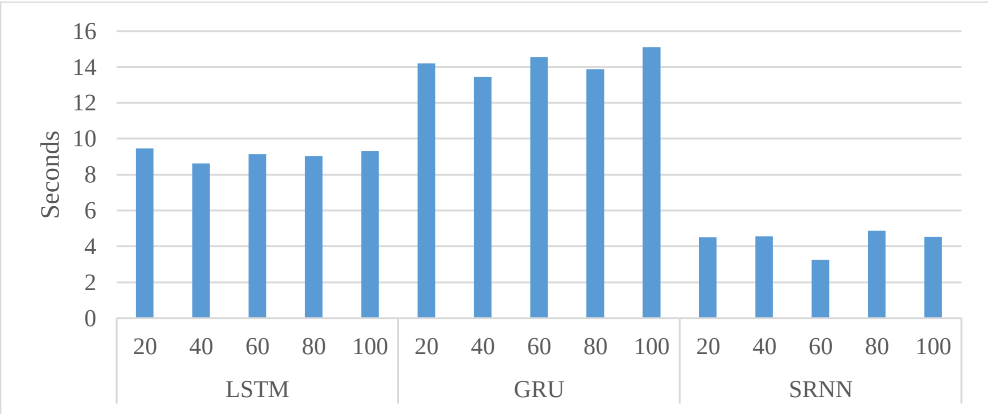

Fig. 3: Computation time of binary classification - weights between 0 and 1

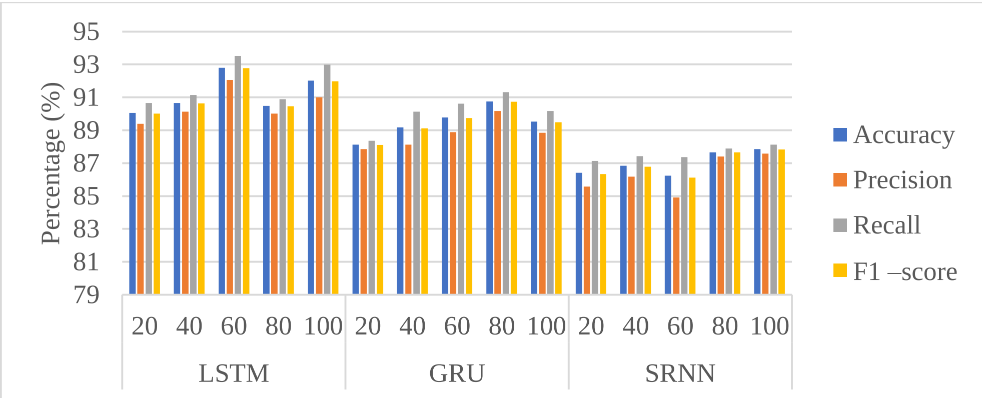

Fig. 4: Performance results of binary classification - weighs between -1 and 1

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

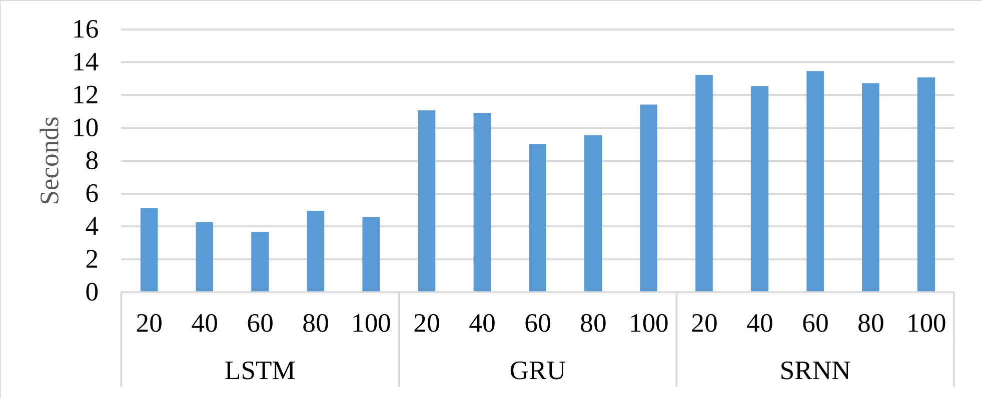

Fig. 5: Computation time of binary classification - weights between -1 and 1

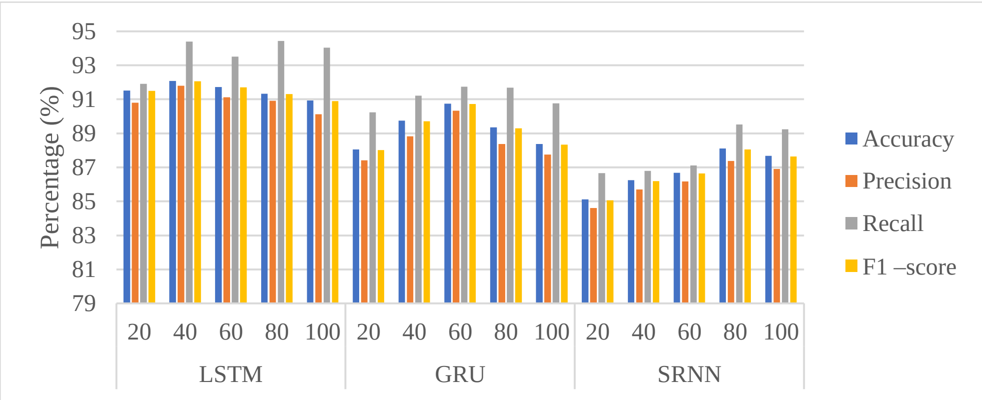

Fig. 6: Performance results of binary classification – weight between -0.5 and 0.5

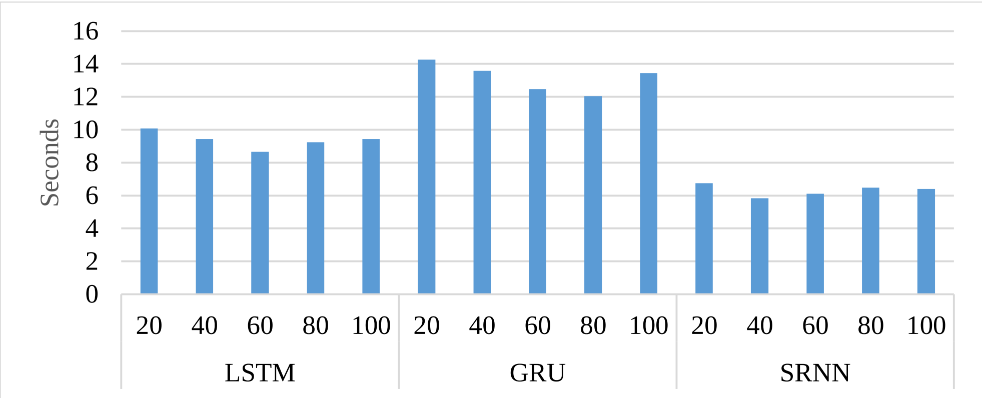

Fig. 7: Computation time of binary classification - weights between -0.5 and 0.5

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

b) *Multi-classification*

Students are divided into three or more performance levels or categories using multiclass categorization based on their CGPA. Deeper insights are obtained by multiclass classification, allowing for customized tactics for varying performance levels. Tables 6,7 and 8 and Figures 8, 10 and 12 show the performance results based on multi-classification and Figures 14,16, and 18 show the computational weights based on different weight ranges and hidden neurons in the hidden layers for LSTM, SRNN, and GRU, respectively. According to the multi-classification, the GRU produced optimal results such as an accuracy of 92.53% when choosing 60 hidden neurons with weights ranging between 0 and 1, an accuracy of 92.72 % when choosing 40 hidden neurons with weights between -1 and 1, and 92.65 % when choosing 40 hidden neurons weights between -0.5 and 0.5. However, SRNN is taking a low computational time compared to other methods such as LSTM and GRU. According to the above result, GRU produced high accuracy with 60 hidden neurons and weights ranging between 0 and 1.

Table 6. Performance results of multiclassification - Weights range between 0 and 1

Methods	HS	Accuracy	Precision	Recall	F1 –score	Time (Sec)
GRU	20	90.55	90.33	90.76	90.54	12.98
	40	90.25	89.83	90.65	90.23	13.44
	60	**92.53**	**91.21**	**93.84**	**92.5**	14.17
	80	91.84	90.65	92.99	91.81	12.2
	100	91.4	90.47	92.29	91.37	15.28
LSTM	20	88.1	86.78	89.28	88.01	7.28
	40	89.01	87.88	90.04	88.95	9.89
	60	89.61	88.37	90.76	89.55	9.08
	80	**89.97**	**88.59**	**91.27**	**89.91**	**8.25**
	100	89.54	88.23	90.76	89.48	8.88
SRNN	20	85.02	84.03	85.85	84.93	3.84
	40	85.91	84.79	86.85	85.81	4.71
	60	86.86	85.54	88.01	86.76	**3.78**
	80	87.48	86.55	88.3	87.41	4.25
	100	**88.13**	**87.35**	**88.82**	**88.08**	4.69

Table 7. Performance results of multiclassification - Weights range between -1 and 1

Methods	HS	Accuracy	Precision	Recall	F1 –score	Time (Sec)
GRU	20	90.59	90.09	91.06	90.57	12.56
	40	91.02	90.58	91.43	91	12.97
	60	**92.72**	**92.17**	**93.26**	**92.71**	13.41
	80	91.68	91.14	92.19	91.66	13.22
	100	91.29	90.73	91.83	91.27	13.13
LSTM	20	89.21	88.32	90.01	89.16	9.24
	40	89.61	88.64	90.51	89.57	9.83
	60	**90.44**	**89.27**	**91.55**	**90.39**	8.65
	80	90.18	89.07	91.22	90.13	8.57
	100	89.48	88.88	90.04	89.45	9.28
SRNN	20	85.13	84.14	85.96	85.04	4.97
	40	86.4	85.59	87.1	86.34	5.11
	60	85.93	85.14	86.59	85.86	5.1

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

	80	**88.32**	**87.35**	**89.19**	**88.26**	6.2
	100	87.9	86.88	88.8	87.83	6.07

Table 8 . Performance results of multiclassification - Weights range between -0.5 to 0.5

Methods	HS	Accuracy	Precision	Recall	F1 –score	Time (Sec)
GRU	20	91.5	90.67	92.29	91.47	12.41
	40	**92.65**	**91.91**	**93.37**	**92.63**	12.64
	60	92.08	91.33	92.8	92.06	13.05
	80	91.21	90.62	91.77	91.19	**11.42**
	100	90.86	90.41	91.29	90.85	13.24
LSTM	20	89.83	88.85	90.73	89.78	8.41
	40	89.61	88.6	90.54	89.56	8.57
	60	**90.72**	**90.12**	**91.28**	**90.7**	**8.82**
	80	90.49	89.89	91.05	90.47	9.23
	100	90.27	89.75	90.75	90.24	9.11
SRNN	20	87.27	86.52	87.92	87.21	4.82
	40	87.78	86.88	88.57	87.72	5.08
	60	88.93	87.75	90.02	88.87	5.01
	80	**89.34**	**88**	**90.57**	**89.27**	4.57
	100	88.83	87.55	90.01	88.76	4.63

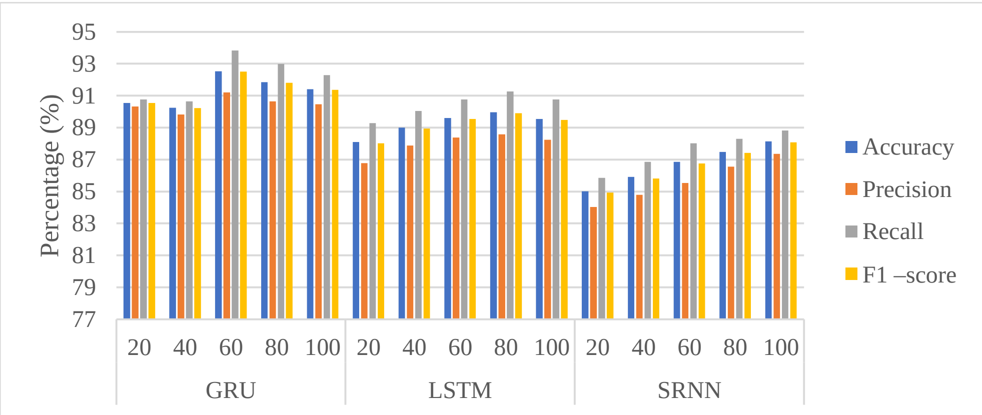

Fig. 8 : Performance results of multi-classification – weight between 0 and 1

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

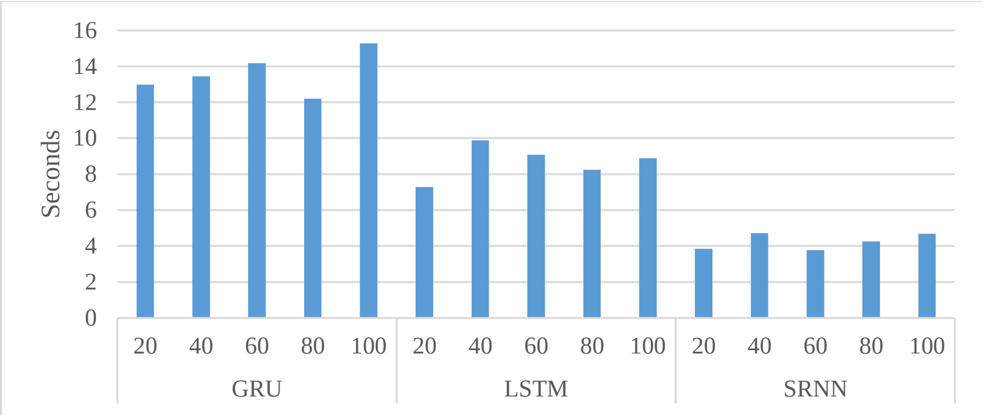

Fig. 9 : Computation time of multi-classification - weights between 0 and 1

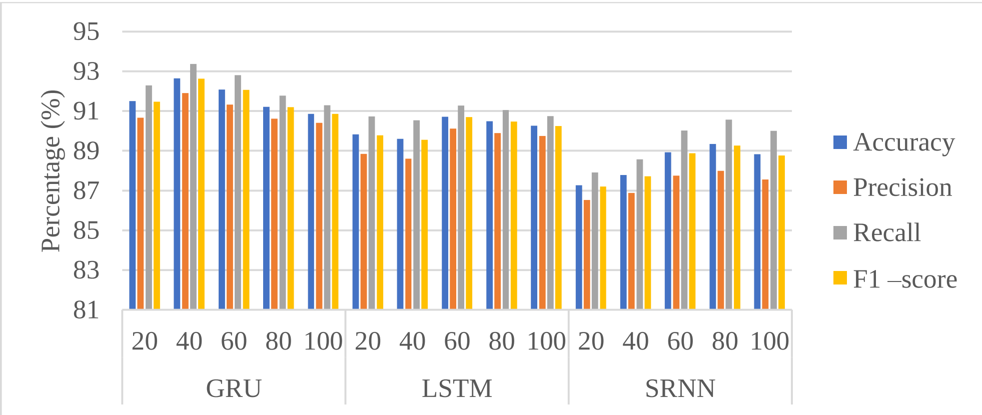

Fig. 10 : Performance results of multi-classification – weight between -0.5 and +0.5

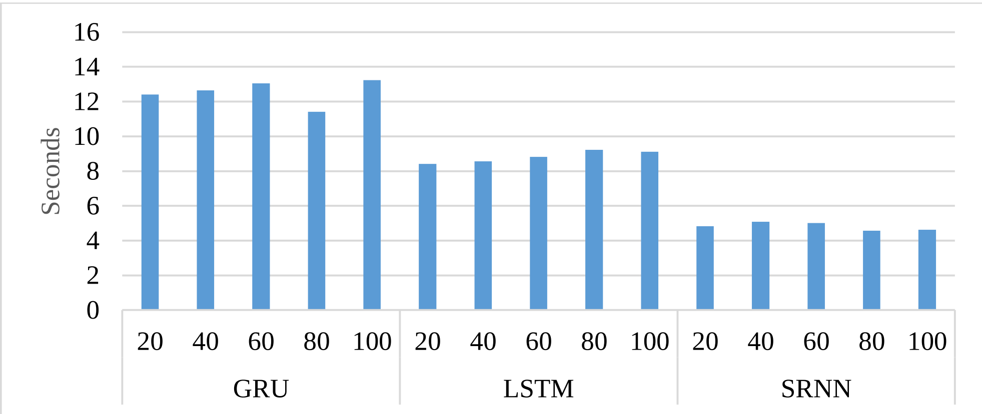

Fig. 11 : Computation time of multi-classification - weights between -0.5 and +0.5

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

Overall experimental results show that SRNN produced low computational time compared with LSTM and GRU. However, GRU produces high accuracy when solving multi-classification problems and LSTM produces high accuracy when solving the problem of binary classification problems.

V. CONCLUSIONS

In this study we compare the performance of SRNN, LSTM and GRU models for prediction of student mental health from social media data. The results show that LSTM gives the best accuracy in prediction due to its ability to effectively capture long-term sequential dependencies, while GRU provides a good compromise between accuracy and computational efficiency in real-time applications. However, SRNN does worse in this respect, as it cannot learn long-term patterns well. These results prove the potential of LSTM and GRU based models for early detection and intervention for mental health in educational institutions. Further research will focus on the addition of more mental health indicators, the creation of hybrid attention-based models, and the implementation of the developed models in real-world web and mobile applications for continuous mental health monitoring.

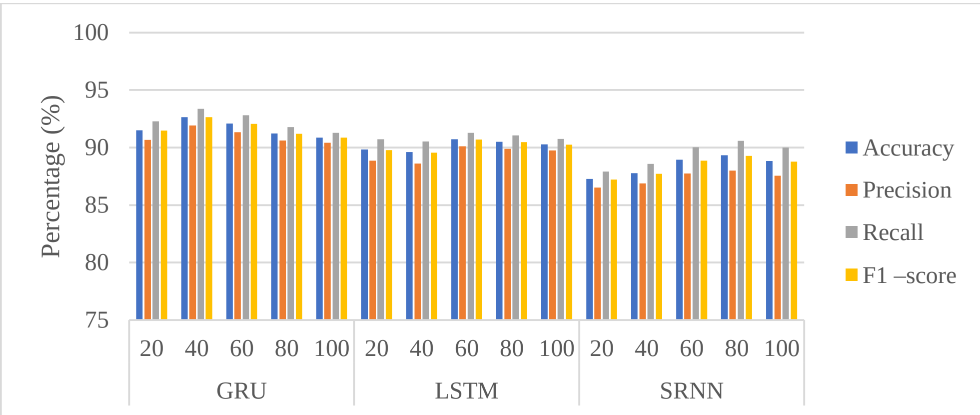

Fig. 12 : Performance results of multi-classification – weight between -1 and 1

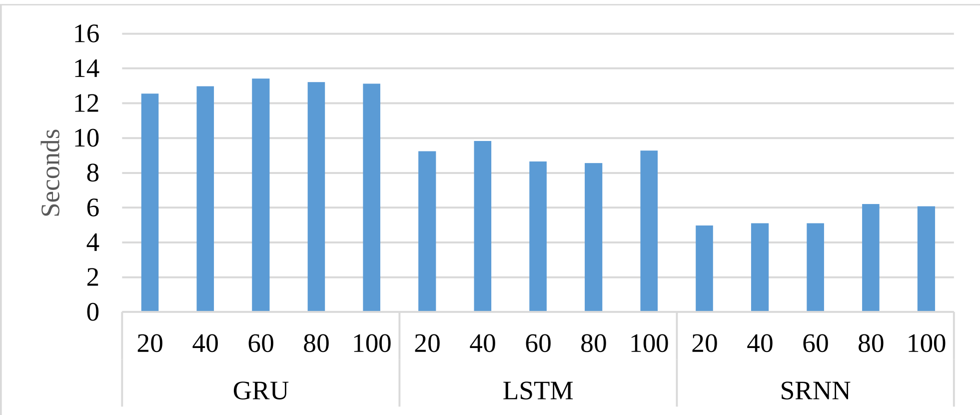

Fig. 13 : Computation time of multi-classification - weights between -1 and 1

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)

ISSN: 2320-088X
Impact Factor: 7.056

References

[1]. M. S. H. Mukta, S. Islam, S. Shatabda, M. E. Ali, and A. Zaman, "Predicting academic performance: Analysis of students' mental health condition from social media interactions," *Behavioral Sciences,* vol. 12, no. 4, p. 87, 2022.

[2]. A. H. Alsabhan *et al.*, "SailMutLoc: a sailfish-mutation enhanced optimization algorithm for student performance prediction," *Ain Shams Engineering Journal,* vol. 17, no. 1, p. 103898, 2026.

[3]. B. Venkatachalam and K. Sivanraju, "Predicting Student Performance Using Mental Health and Linguistic Attributes with Deep Learning," *Revue d'Intelligence Artificielle,* vol. 37, no. 4, 2023.

[4]. E. Haghish, M. Obaidi, T. Strømme, T. Bjørgo, and C. Grønnerød, "Mental health, well-being, and adolescent extremism: a machine learning study on risk and protective factors," *Research on Child and Adolescent Psychopathology,* vol. 51, no. 11, pp. 1699-1714, 2023.

[5]. C. Du, C. Liu, P. Balamurugan, and P. Selvaraj, "Deep learning-based mental health monitoring scheme for college students using convolutional neural network," *International Journal on Artificial Intelligence Tools,* vol. 30, no. 06n08, p. 2140014, 2021.

[6]. K. Vijayakumari, "Artificial Intelligence in Education: A Systematic Literature Review of Trends, Applications, and Challenges," *Advances in Computer Engineering and Science,* vol. 1, no. 1, pp. 1-5, 2025.

[7]. M. Jibril, "Machine Learning for Predicting Students' Academic Performance in Selected Educational Technology Courses in Nigerian Universities," *Journal of Educational Technology Development and Exchange (JETDE),* vol. 19, no. 2, pp. 20-37, 2026.

[8]. T. Bharathi and R. Manavalan, "Data-Driven Student Performance Prediction: A Hybrid Approach Using Weighted Extreme Learning Machine and Coati Optimization Algorithms," *SN Computer Science,* vol. 7, no. 1, p. 72, 2026.

[9]. J. Li *et al.*, "The multi-parameter optimized belief rule base for predicting student performance with interpretability," *Scientific Reports,* 2026.

[10]. C.-A. Buzducea *et al.*, "Machine Learning in Education: Predicting Student Performance and Guiding Institutional Decisions," *Education Sciences,* vol. 16, no. 1, p. 76, 2026.

[11]. J. Wang and Y. Yu, "Machine learning approach to student performance prediction of online learning," *PloS one,* vol. 20, no. 1, p. e0299018, 2025.

[12]. R. Arun, M. Vanithamani, S. Natarajan, K. K. Thoti, S. S. Bharathi, and P. Raju, "The impact of social networking sites on students' social wellbeing and academic performance based on RBF-Elman techniques," in *Hybrid and Advanced Technologies*: CRC Press, 2025, pp. 125-130.

[13]. K. Li, "Using biosensors and machine learning algorithms to analyse the influencing factors of study tours on students' mental health," *Molecular & Cellular Biomechanics,* vol. 21, no. 1, pp. 328-328, 2024.

[14]. Y. Ding, X. Chen, Q. Fu, and S. Zhong, "A depression recognition method for college students using deep integrated support vector algorithm," *IEEE access,* vol. 8, pp. 75616-75629, 2020.

[15]. A. S. Aljaloud *et al.*, "A deep learning model to predict Student learning outcomes in LMS using CNN and LSTM," *IEEE Access,* vol. 10, pp. 85255-85265, 2022.

[16]. V. Das Swain, J. Ye, S. K. Ramesh, A. Mondal, G. D. Abowd, and M. De Choudhury, "Leveraging Social Media to Predict COVID-19–Induced Disruptions to Mental Well-Being Among University Students: Modeling Study," *JMIR Formative Research,* vol. 8, p. e52316, 2024.

[17]. I. J. Ratul, M. M. Nishat, F. Faisal, S. Sultana, A. Ahmed, and M. A. Al Mamun, "Analyzing perceived psychological and social stress of university students: A machine learning approach," *Heliyon,* vol. 9, no. 6, 2023.

[18]. Q. Chen and B. G. Lee, "Deep learning models for stress analysis in university students: a sudoku-based study," *Sensors,* vol. 23, no. 13, p. 6099, 2023.

[19]. S. Pandey and S. Sharma, "A comparative study of retrieval-based and generative-based chatbots using deep learning and machine learning," *Healthcare Analytics,* vol. 3, p. 100198, 2023.

[20]. N. Kirlic *et al.*, "A machine learning analysis of risk and protective factors of suicidal thoughts and behaviors in college students," *Journal of American college health,* vol. 71, no. 6, pp. 1863-1872, 2023.

[21]. L. Chen, "Exploring the reform of college mental health education based on multi-source data," *Applied Mathematics and Nonlinear Sciences,* 2024.

International Journal of Computer Science and Mobile Computing
A Monthly Journal of Computer Science and Information Technology

S. Anitha, Volume 15, Issue 7, July 2026, pg. 66-84
(An Open Accessible, ISI Indexed, Fully Refereed and Peer Reviewed Journal)
ISSN: 2320-088X
Impact Factor: 7.056

[22]. Q. Gao and Y. Wei, "Understanding the cultivation mechanism for mental health education of college students in campus culture construction from the perspective of deep learning," *Current Psychology,* vol. 43, no. 2, pp. 1715-1732, 2024.
[23]. J. Luo and S. Deng, "Evaluation Model of the Mental Health Education Effectiveness Based on Deep Neural Networks," *Journal of computing and information technology,* vol. 31, no. 1, pp. 57-72, 2023.
[24]. Y. Jia, "Impact of Music Teaching on Student Mental Health Using IoT, Recurrent Neural Networks, and Big Data Analytics," *Mobile Networks and Applications,* pp. 1-20, 2024.
[25]. S. Hochreiter, "Long Short-term Memory," *Neural Computation MIT-Press,* 1997.
[26]. R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, 2017: IEEE, pp. 1597-1600.
[27]. H. Liu, F. Hussain, C. L. Tan, and M. Dash, "Discretization: An enabling technique," *Data mining and knowledge discovery,* vol. 6, pp. 393-423, 2002.
[28]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.
[29]. D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing,* vol. 97, p. 105524, 2020.
[30]. T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern recognition,* vol. 48, no. 9, pp. 2839-2846, 2015.
[31]. A. Bashaiwth, H. Binsalleeh, and B. AsSadhan, "An Explanation of the LSTM Model Used for DDoS Attacks Classification," *Applied Sciences,* vol. 13, no. 15, p. 8820, 2023.

