Comparing CNN Regression with KNN and RF Algorithms in Prediction of Student Need for Support

Ahmed M. D. E. Hassanein,
Systems and Information Department,
Engineering and Renewable Energy Research Institute, National Research Centre (NRC), Dokki, Giza, Egypt. Postal Code: 12622.
Email: ahmed.diaa.hassanein@gmail.com

Al-Shaimaa Gamal Ramadan,

Omar Tarek Ibrahim,
Electrical Communication and Electronics Systems Dept., School of Engineering, October University for Modern Sciences and Arts (MSA), Al-Wahat Road, Giza, Egypt.
Email: alshaimaa.gamal@msa.edu.eg, omar.tarek20@msa.edu.eg

***Abstract*— Educational institutions are shifting towards increasing their dependence on online courses due to the recent challenges. Analytic tools are becoming increasingly important for online courses to help teachers monitor the progress of their students. In this study, we focus on the element of student progress throughout the time of the semester and redefining the predicted class to achieve higher accuracy. The possibility of predicting the student need for help as the marks of the student accumulate throughout the semester is explored. The methods of prediction which are used are namely the correlation method and two machine learning methods. The Random Forest (RF) and K Nearest Neighbor (KNN) algorithms are investigated. The accuracy in predicting student need for help is 79% for the KNN method and 86% for the RF method. The results come in confirmation with previous research which stated that the RF algorithm is better than other algorithms in predicting student grades. Moreover, the achieved percentage of accuracy which is 86% for the RF algorithm is higher than that achieved in previous research. This is due to redefining the class, which is predicted, to be the degree of need for support/help rather than the actual grade of the student. In addition, the time factor of mark progress as the semester activities increase is taken into consideration. Then, a four layer CNN regression network is proposed to predict the class of student need for help. Two types of datasets are used. The first is two class dataset which predicts either the student need help or not. The second is five class dataset which predicts degrees of the student need for help. For both datasets, the accuracy of prediction for the CNN regression network reached hundred percent with a RMSE of 5.**

***Keywords*— KNN algorithm; RF algorithm; Grade prediction; Correlation; Prediction Accuracy.**

I. INTRODUCTION

Educational institutions have been passing through a lot of changes due to the major problems that is facing the world. Covid19 has forced all daily routines in human lives to change among which the educational systems. Moreover, the increase in number of students who are interested in receiving education has encouraged higher educational institutions to increase the number of online courses to enable students to access course material at any time. In the following, the published work in this context which is performed by previous researchers is described.

Due to the outspread of covid-19, many academic institutions changed the method of delivering courses from being face to face to being a hybrid one or even online only [1]. The hybrid method is a one which blends between face to face teaching and online delivery of a course. The online teaching methods requires more statistical analysis of the behavior of students on the online platforms combined with the student grades to predict the final grades of the students. Badal et al. propose a predictive model that takes into account student activities and grades to make accurate predictions about the final grades of the students [1], [2]. The Random Forest algorithm proved to be the best in predicting the grades of the students based on the proposed predictive model [1].

Students from all backgrounds have become able to join educational institutions. Students' needs are different according to their physics, social and economic abilities. Educational institutions have become under increasing demand to identify students who need help as soon as possible so that they would be able to give them the help they need in real time. Instructors used to predict performance of students in their courses manually through averaging their marks and personally following the activities of the students [3]. In today's advancement of data science, machine learning can be used to enable instructors to identify students with problems. Anderson et al. proposes comparing the performance of the support vector machine computational method and the simple average manual method to predict final grades of the students [4]. Both methods have produced almost the same error margin in predicting the final grades [4]. The support vector machine was computationally very demanding method over the simple average one [5]. The paper concludes that a lot of work needs to be done to use machine learning algorithms in students' grades prediction [5].

Teachers have to widen their use of the online methods of course delivery and student evaluation. Programmers have to make use of the analytics tools found in the online course delivery methods to provide all interested parties with performance monitoring reports. Students registered in courses which are taught through online platforms always struggle to get the help they need especially that the face to face interaction is not available most of the time. Analytics tools can be of great help in such cases. Instructors as well as students can use them to foresee the performance of each student based on the activities and grades of each student [6]. Accordingly, the kind of help required for each student can be accurately defined and delivered in the appropriate time to the student. In the following lines, we lay out five examples to show the efforts of programmers and researchers.

In the first example a course, in which the marks of 1282 registered students, is used to test two issues. First the performance of six algorithms is to be analyzed to find which best can be used [7]. The six are K-Nearest Neighbor (kNN), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM) [8], Naïve Bayes (NB), and Decision Tree (J48) [7]. Second, Bujang et al. proposed a multiclass prediction model to decrease the error in predicting final grades of students [7], [9].

In the second example, Iqbal et al. uses the Restricted Boltzmann Machines (RBM), Collaborative Filtering (CF), and Matrix Factorization (MF) techniques to analyze the data of student academic activities in the Engineering department of one of the universities in Pakistan [6]. The RBM was found to be the best technique in predicting the performance of the students [6].

In the third example, Venkat et al. analyzed the performance of several machine learning techniques on several features per each student to predict his/her grade [10], [11]. The features are many including CGPA, attendance, several exams grades and some biometric data [10]. Principal Component Analysis is used to reduce the dimensions of the dataset used but results have shown that it is not suitable for this aim [10], [8].

In the fourth example, an accurate predictive model which can be used to predict with high accuracy the final grade of a student is found to be in high demand [12]. The need for such a model is increasing because of the huge amount of data which is generated every day by the online platforms. The data for the 10th, 12th and other semester grades are collected and a predictive model is proposed based upon them [13]. The performance of the proposed model is evaluated using the KNN, Decision tree, Binomial logical regression, and Entropy classifiers [13].

In the fifth example, Prakash et al. proposes a model that is based on the previous grades of students and some other socio-economic factors [14], [15]. The machine learning model helps academic institutions predict future final grades of students in their courses. The model helps teachers to identify students who need help to improve their final grades [16]. In addition, the model helps students to foresee their predicted grades as a warning to ask for help or work on their problems to improve their marks [16].

In this paper, we discuss the possibility of prediction of student need for help. Predicting the student degree of need for help can be of great help to the teacher [17]. This enables the teacher to see the progress of students and identify potential problems and students who suffer academically to take proper action in the right time [17]. The action can be giving extra readings or problems to clarify fundamental concepts [17].

In section two, the characteristics of the features in the collected dataset are discussed. The time plan at which the features/activities take place is illustrated. In section three, the preprocessing operations that are applied on the dataset are discussed. Other extra features are calculated in section three. The averages for all features are analyzed. In section four, the mathematical background for our calculation is illustrated. In section five, the results of our calculations are described. The correlation between features and the prediction of student class are illustrated. A comparison between the class prediction using the CNN Regression and KNN and RF algorithms is discussed. Finally, the conclusion is drawn is section six.

II. Dataset

To solve the problem of predicting a student degree of need for help, we collect a dataset of the marks of student throughout the semester in one of the basic courses taught to engineering and computer science students in several universities. The marks are for all activities given to students throughout the semester. In this section, the characteristics of the dataset to be collected, the time plan, and the method of collection are explained.

A. Characteristics

As shown in table 1, the collected data is presented here in the form of average percentages and standard deviation to represent the range of marks of the students.

TABLE I. THE COLLECTED FEATURES USED IN THIS PAPER ARE SHOWN WITH THE AVERAGE 'AVG' AND STANDARD DEVIATION 'SD'.

#	Type	Feature	Avg %	SD(+/-)
1	Quiz	Q1	85.3	20.2
2		Q2	59.1	24.5
3		Q3	81.5	21.2
4	Assignment	Ass1	72.9	31.4
5		Ass2	67.2	33.2
6		Ass3	84.4	31.4
7	Attendance	Att.	70.5	26.9
8	Final Lab	FLab	88.3	17.2
9	Lab Report	LabR	93.3	17.5
10	Mid Term	MT	69.7	21

In table 1, the features representing all activities given to students are listed. The activities are mainly divided into quizzes given the symbol 'Q', assignments given the symbol 'Ass' and laboratory work given the symbol 'Lab'. Three quizzes are given to the students throughout the semester which are 'Q1', 'Q2' and 'Q3'. Three assignments are given to the students throughout the semester which are 'Ass1', 'Ass2' and 'Ass3'. The laboratory work includes two marks. The 'FLab' stands for the mark of the final exam of the laboratory. The 'LabR' stands for the mark of the weekly experimental work performed by the student inside the laboratory. 'Att' stands for the mark of the attendance of each student. The 'MT' stands for the mark of the midterm exam. The total number of attributes or students for which these features are collected is one hundred and sixty students.

B. *Time Plan*

The time plan and method by which each of the features is gathered are discussed.

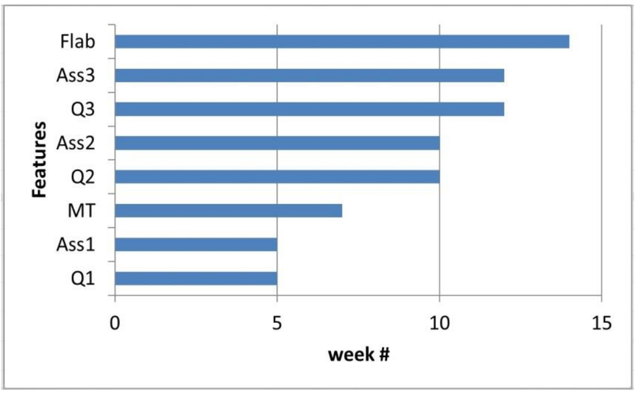

Fig. 1. *The time line for executing the activities of the course throughout the semester is plotted.*

The time plan of carrying out each one of those activities which are mentioned in table 1 is represented in fig. 1. The 'Q1', 'Q2' and 'Q3' quizzes are performed in week 5, week 10 and week 12 respectively. The 'Ass1', 'Ass2' and 'Ass3' assignments are performed in week 5, week 10 and week 12 respectively. The 'MT' exam and the 'FLab' exam are performed in week 7 and week 14 respectively.We divide the semester into landmark weeks at which we can calculate the success of our classification algorithms. The landmark weeks are 5, 7, 10, 12 and 14.

The marks of the 'Att' and laboratory reports 'LabR' are accumulated throughout the whole semester. The marks of all quizzes 'Q' and assignments 'Ass' are processed once at the time of submission of each quiz or assignment. The marks of the 'FLab' and 'MT' are processed once at the time of performing the final lab exam and midterm exam respectively.

III. PREPROCESSING OPERATIONS

The method by which the marks of the students are calculated to form the final grade of the student is discussed. According to the final grade, prediction of the degree by which a student needs help is defined. The software which is used to perform our calculations is listed.

A. *Calculated Features*

In this subsection, the method by which the individual marks of each student are calculated is shown.

TABLE II. THE CALCULATED FEATURES USED IN THIS PAPER ARE SHOWN WITH THE AVERAGE 'AVG' AND STANDARD DEVIATION 'SD'.

#	Type	Feature	Avg %	SD(+/-)
1	Best2Q	B2Q	85.6	14.3
2	Best2Ass	B2Ass	85	20.2
3	Total Lab	TLab	91.6	16.3
4	Course Work	CW	89.2	14.8

As shown in table 2, the 'B2Q', 'B2Ass', 'TLab' and 'CW' features are calculated as explained in the following lines. At the end of semester, the 'B2Q' which stands for the best 2 quizzes is calculated for each student to be included in the final grade. The 'B2Ass' stands for the best 2 assignments. It is calculated for each student to be included in the final mark. The 'TLab' stands for the total laboratory mark which is equal to the sum of the 'FLab' and 'LabR' marks. The total course work 'CW' for each student is calculated by summing the marks of 'B2Q', 'B2Ass', 'Att' and 'TLab'. The 'CW' mark and the 'MT' mark form the final mark of a student.

The average and standard deviation of the 'B2Q' marks are 85.6% and 14.3 respectively. As for the 'B2Ass' marks, the average and standard deviation are 85% and 20.2 respectively. The average and standard deviation of the 'TLab' marks are 91.6% and 16.3 respectively. As for the 'CW' marks, the average and standard deviation are 89.2% and 14.8 respectively.

B. *Software*

We have used the WEKA [18] software and Excel sheets [19] software to make the calculations for our predictions and classification results and displaying our outputs. In our calculations, we have distributed our data to have 70% of its volume be used in training and

30% of its volume be used in testing. Ten-fold cross validation is applied. As for our algorithms, detailed information and background about the KNN and RF algorithms can be found in references [7] and [10].

In our dataset, we have two classes which are ‘Need help’ and ‘Don’t need help’. The ‘Need help’ cases are those students who are expected to receive final mark below 70%. But, the ‘Don’t need help’ cases are those students who are expected to receive final grade above 70%. For the ‘Need help’ cases the class is made equal to 0 while for the ‘Don’t need help’ cases the class is made equal to 1. The total number of features is fourteen, four of them are calculated as defined in section three and ten are collected from the student activities as defined in section two. The dataset is the collection and calculation of the percentages of marks which are collected from the students.

IV. EQUATIONS

In this section, the formulas used in obtaining our results are illustrated. The Standard deviation is used to give an overall view of the dataset. The standard deviation equation is [20]:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \tag{1}$$

Where σ is the standard deviation, μ is the average of the feature under consideration, x_i is the values of the feature and N is the total number of values for this feature.

The Average is used to give an overall view of the dataset as well. The average equation is [20]:

$$A = \frac{1}{N} \sum x_i \tag{2}$$

Where A is the average, x_i is the values of the feature and N is the total number of values for this feature.

The Correlation is used to find out how the different features are related to each other. The correlation equation is [20]:

$$r = \frac{N\left(\sum x_i y_i\right) + \left(\sum x_i \sum y_i\right)}{\sqrt{\left(N \sum x_i^{\,2} - \left(\sum x_i\right)^2\right)\left(N \sum y_i^{\,2} - \left(\sum y_i\right)^2\right)}} \tag{3}$$

Where r is the correlation value, x_i is the values of one feature, y_i is the values of another feature and N is the total number of values for one feature.

The Convolution is used to find out the convolution between two matrices. Signal x can be convoluted by another signal y by using the equation [21]:

$$C(j,k) = \sum_p \sum_q x(p,q) y(j-p+1, k-q+1) \tag{4}$$

Where C is the convolution matrix, p and q are indexes for the two signals $x(p,q)$ and $y(j-p+1, k-q+1)$.

The regression equation is used to predict a range of continuing classes. The Regression equation is [21]:

$$y = a + rx \tag{5}$$

Where y is the prediction, x is the input signal while a and r are parameters to modify future predictions. The equation for a is [21]:

$$a = \frac{\sum y - r \sum x}{N} \tag{6}$$

Where N is the total number of values for this feature. The r parameter is the correlation coefficient described in eq. (3).

V. RESULTS AND DISCUSSION

In this section, three methods of classifying our dataset are discussed. We start by calculating the convolution between the different features to find out how they are related. Next, the KNN and RF algorithms are applied to the dataset to classify it into two classes. Then, the CNN regression algorithm is applied to classify the dataset into two classes in one trial. In another trial, the dataset is classified into five classes.

A. *Correlation*

In this subsection, we discuss the results obtained when calculating the correlation between each of the fourteen features and some selected features. As shown in table 3, we have eliminated from the correlation matrix the features which are not logically related to each other relative to our study here. For example, finding the correlation between any of the quizzes and the laboratory or assignment features is meaningless as they are not related to them in this study. But, finding the correlation between ‘B2Q’ and all fourteen features can help us in highlighting which features affects the value of the ‘B2Q’ feature. In table 3, we are interested in discussing the features whose correlation value exceed 0.7 which reflects that the features are highly correlated.

TABLE III. THE CORRELATIONS BETWEEN THE ACTIVITIES OR FEATURES WHICH ARE USED IN THIS PAPER ARE CALCULATED.

	class	MT	CW	TLab	Att	B2Ass	B2Q
Q1	0.20	0.25	0.56	0.46	0.35	0.31	0.78
Q2	0.51	0.41	0.56	0.39	0.48	0.55	0.52
Q3	0.31	0.26	0.55	0.39	0.41	0.37	0.73
B2Q	0.38	0.39	0.77	0.57	0.50	0.51	1.00
Ass1	0.51	0.57	0.60	0.40	0.49	0.77	0.34
Ass2	0.43	0.35	0.54	0.36	0.45	0.65	0.37
Ass3	0.36	0.39	0.73	0.60	0.59	0.74	0.48
B2Ass	0.49	0.55	0.88	0.66	0.65	1.00	
Att	0.49	0.47	0.70	0.56	1.00		
FLab	0.35	0.43	0.73	0.87	0.47	0.51	0.47
LabR	0.29	0.43	0.88	0.97	0.55	0.67	0.57
TLab	0.33	0.46	0.89	1.00			
CW	0.47	0.55	1.00				
MT	0.65	1.00					
class	1.00						

In table 3, 'B2Q' is highly correlated with 'Q1' and 'Q3' with correlation values of 0.78 and 0.73 respectively. For most students, 'Q1' and 'Q3' are the ones to be taken in calculating the mark of the best two quizzes 'B2Q'. In table 1, we can see that for the quizzes the first and third quizzes have the highest averages. 'B2Ass' is highly correlated with 'Ass1' and 'Ass3' with correlation values of 0.77 and 0.74 respectively. We can see that 'Ass1' and 'Ass3' are the ones to be taken in calculating the mark of the best two assignments 'B2Ass'. For the assignments in table 1, we can see that 'Ass1' and 'Ass3' have the highest averages. The 'FLab' and 'LabR' have the highest averages in all features. The 'TLab' is highly correlated with 'FLab' and 'LabR' with correlation values of 0.87 and 0.97 respectively. The 'TLab' is equal to the sum of 'FLab' and 'LabR'. The 'FLab' and 'LabR' are targeting a different element in education which is applied and practical work. The 'Att' is almost equally correlated with all features with none of the features being highly correlated with it. Attendance is important for all activities and features. The 'CW' is highly correlated with 'B2Q', 'B2Ass', 'Att' and 'TLab' with correlation values of 0.77, 0.88, 0.7 and 0.89 respectively. The 'CW' targets activities which include group work and theoretical basis. The 'CW' is also highly correlated with 'Ass3', 'FLab' and 'LabR' with correlation values of 0.73, 0.73 and 0.88 respectively. 'Ass3', 'FLab' and 'LabR' are included and embedded in other marks which are 'B2Ass' and 'TLab' as explained before. Both the 'MT' and class values are not highly correlated with any of the fourteen features. The 'MT' targets activities which include individual work to solve, analyze and memorize on one's own self which is completely different from other activities. The low correlation between class and all other fourteen features means that no features can be considered as an indicator to the predicted class. Time progress of marks is not taken into consideration to predict the degree of help for each student when calculating the correlation. Next, we discuss the way by which the progress in the marks of the students, as the weeks of the semester pass on, increases the ability of our selected algorithms in predicting the class of the student.

B. KNN and RF Algorithms

As discussed before, the activities/features done throughout the semester can be allocated in certain milestones in terms of timing. So, in week 5, three activities are included in predicting the class of a student which are 'Q1', 'Ass1' and 'Att'. In week 7, four activities are included in our prediction calculations. The 'MT' activity and the three activities in week 5 are included in our calculations. In week 10, six activities are included in our calculations. 'Q2' and 'Ass2' in addition to the four in week 7 are included in our calculations. In week 12, eight activities are included in our calculations. 'Q3' and 'Ass3' in addition to the six in week 10 are included in our calculations. In week 14, nine activities are included in our calculations. 'LabR' activity in addition to the eight activities in week 12 is included in our calculations. At the end of semester, all fourteen activities mentioned in table 1 and 2 are included in our predictions. The results of applying the two algorithms namely KNN and RF are discussed.

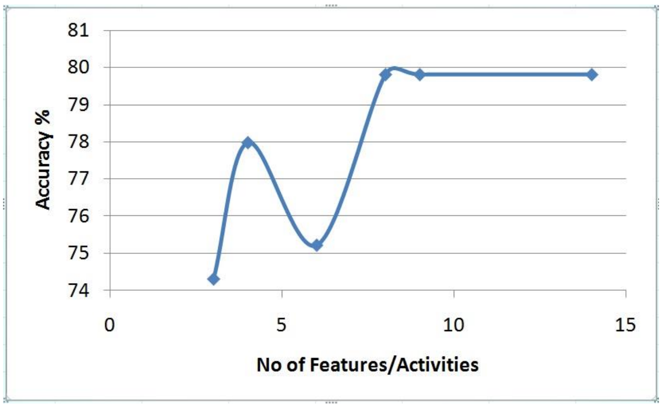

Fig. 2. *Using KNN, accuracy of predicting the class of help needed by the student as the number of activities/features increase throughout the semester is shown.*

As shown in fig. 2, the accuracy of predicting the class of each student is calculated using the KNN algorithm. The class represent whether the student needs help or not. The percentage of accuracy increases as the number of features/activities increase or in other words as the weeks of the semester passes by. As seen in fig. 2, the percentage of accuracy increases as the features increase from three up to eight with one exception. The percentage accuracy reaches almost 79.8% at eight features. But at six features, the accuracy drops to 75.5%. This can be due to the inclusion of 'Q2' and 'Ass2' which have the lowest percentages among all other quizzes and assignments as shown in table 1. Then in fig 2, the percentage accuracy saturates at the value of 79.8% as the features increases from eight to fourteen.

As shown in fig. 3, the accuracy of predicting the class of each student is calculated using the RF algorithm. The percentage of accuracy increase as the number of features/activities increase or in other words as the weeks of the semester passes by. As seen in fig. 3, the percentage of accuracy increases steadily as the features increase from three up to fourteen with one exception. The percentage accuracy reaches almost 84.5% at fourteen features. But at six features, the accuracy drops to 80%. Next, we calculate the percentage accuracy using the KNN and RF algorithms based on 6 features only. The six features are the main ones which compose the total grade of a student which are 'B2Q', 'B2Ass', 'Att', 'TLab', 'MT' and 'CW'. All other features are included in these six features or in other words calculated inside them.

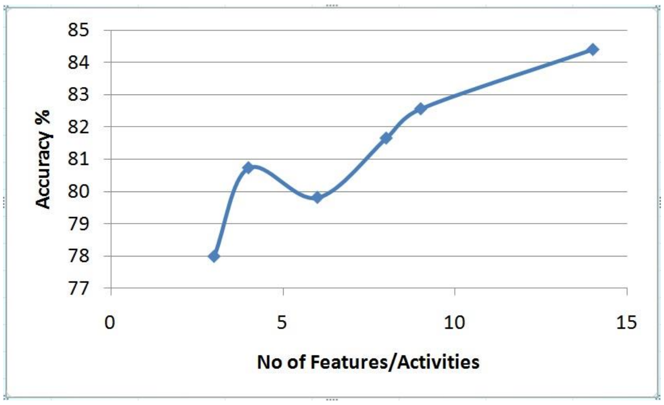

Fig. 3. *Using RF, accuracy of predicting the class of help needed by the student as the number of activities/features increase throughout the semester is shown.*

As shown in fig. 4, the accuracy of predicting the class of each student is compared when using the RF and KNN algorithms. The main six features are used in calculating the accuracy of prediction in both cases. As seen in fig. 4, the percentage of accuracy when using the RF algorithm is higher than that when using the KNN. For the KNN algorithm, the percentage of accuracy is almost 79% when using six features. For the RF algorithm, the percentage of accuracy is almost 86% when using six features.

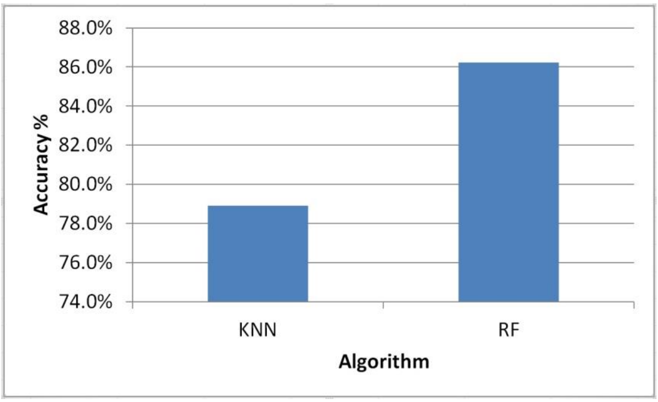

Fig. 4. *Comparing the results of using the RF and KNN algorithms on student class of help using the main 6 features.*

C. CNN Regression

In this subsection, we use a CNN network that is composed of four layers. An input layer that is used to enter the data to the network. A convolutional layer to apply eight times a 2x2 convolution. A Fully Connected layer and regression layer to output the results. The regression layer has the advantage of predicting a continuous range of class not only discrete classes [21]. The equations for some of the layers are mentioned in section four.

We use the Stochastic Gradient Descent with Momentum 'SGDM' optimization algorithm [22]. The maximum number of epochs is 100. The initial learn rate value is 0.001. The learn rate schedule is 'piecewise'. The value of the learn rate drop factor is 0.1. The value of the learn rate drop period is 20. The Root Mean Square Error (RMSE) and Loss parameters [23] are used to evaluate the success achieved in the network. The earlier parameter represents the difference between the output and the expected output [23]. The later parameter represents the rate of change of error at prediction as the number of epochs increase [23].

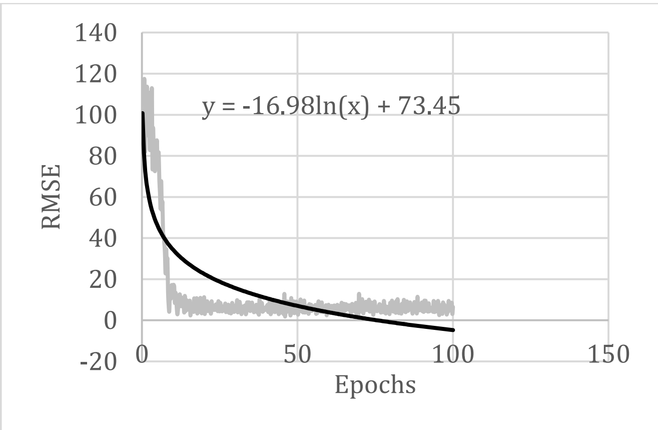

Fig. 5. *The RMSE results and a best fit are plotted for prediction of student class (2 classes) of help using the main 6 features.*

As shown in Fig. 5, the RMSE is plotted as the number of epochs increase. The dataset used contains only two classes either the student need help or not. The equation of the best fit curve is written on the figure as well. The values of the RMSE drops exponentially as the number of epochs increase from a starting value of 73.45. The rate by which the curve drops is -16.98.

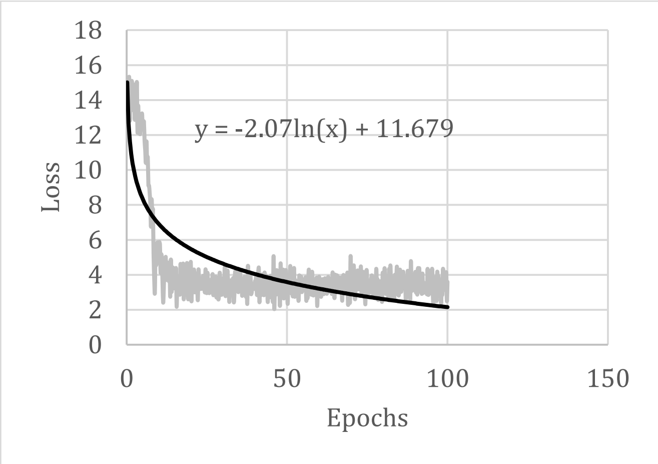

Fig. 6. *The Loss results and a best fit are plotted for prediction of student class (2 classes) of help using the main 6 features.*

As shown in Fig. 6, the Loss is plotted as the number of epochs increase. The dataset used contains only two classes either the student need help or not. The equation of the best fit curve is written on the figure. The values of the Loss drops exponentially as the number of epochs increase from a starting value of 11.67. The rate by which the curve drops is -2.07. The accuracy of prediction reached hundred percent with a RMSE of 5.

Next, we try to make use of the regression layer by increase the range of continuous classes that the network would predict. Now, the dataset is changed to predict 5 classes of student need for support starting from No need is needed and ending by Very Urgent need for the support is needed. The five new categories are 'No help is needed', 'On the verge of needing help', 'help is needed', 'Urgent help is needed' and 'Very Urgent help is needed'. We use the same network with the same values of parameters mentioned above to predict five classes instead of two classes.

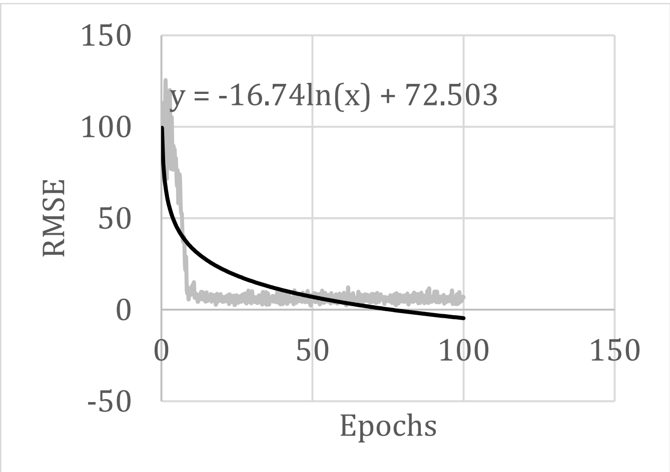

Fig. 7. *The RMSE results and a best fit are plotted for prediction of student class (5 classes) of help using the main 6 features.*

As shown in Fig. 7, the RMSE is plotted as the number of epochs increase. The dataset used contains five classes. The equation of the best fit curve is written on the figure. The values of the RMSE drops exponentially as the number of epochs increase from a starting value of 72.5. The rate by which the curve drops is -16.74.

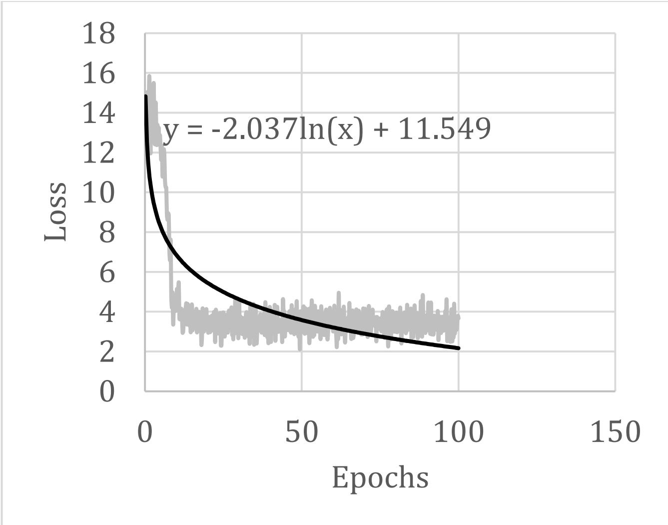

Fig. 8. *The Loss results and a best fit are plotted for prediction of student class (5 classes) of help using the main 6 features.*

As shown in Fig. 8, the Loss is plotted as the number of epochs increase. The dataset used contains five classes. The equation of the best fit curve is written on the figure. The values of the Loss drops exponentially as the number of epochs increase from a starting value of 11.54. The rate by which the curve drops is -2.03. The accuracy of prediction reached hundred percent with a RMSE of 5.

The proposed CNN regression network continues to predict the classes of student need for help even after increasing the number of classes with an almost equal accuracy.

VI. Conclusion

In this work, we study the success of predicting a student degree of need for help. The marks of fourteen features/activities are collected and calculated from students throughout the semester. Then, the possibility of predicting student class is studied by using either the correlation relation or machine learning algorithms. We have shown that the correlation between student class and all other fourteen features is not high. All correlation values between student class and all fourteen features are below 0.7. The correlation relation is not suitable for predicting student class from any of the fourteen features as it is concerned with one to one relation between two ranges of values. The time evolution of mark accumulation for each student as the number of activities increase throughout the semester is not taken into consideration. We continued our work by using the RF and KNN algorithms predict student class. The percentage of accuracy in predicting student class is almost 79% for the KNN algorithm and

86% for the RF algorithm. This result is verified in the research papers mentioned before such as that performed by Badal et al. [1]. Gaftandzhieva et al. show that the Random Forest algorithm is the best in predicting student grades with an accuracy of 78% [24]. The study shows that the Random forest is better than KNN and other algorithms in predicting students who are about to fail the course after eight weeks from start of the semester [24]. In this work, we have also succeeded in increasing the accuracy of prediction by the RF algorithm by an almost 10% to be 86%. This is because we have break down the problem of grade prediction to the problem of predicting the degree of student need for help. This means that we are trying to predict the range of grades into which the marks of the student falls rather than a specific grade for each student. Also, this is because of the inclusion of the element of the progress in the marks of each student that is to say the time evolution of the students' marks. Moreover, a four layer CNN regression network is proposed to predict the class of student need for help. The regression layer gives us the advantage of predicting a larger range of values of classes. Two types of classes are used. The first one is the two class dataset which predicts either the student need help or not. The second one is the five class dataset which predicts degrees of the student need for help. For both datasets, the accuracy of prediction for the CNN regression network reached hundred percent with a RMSE of 5.

References

[1] Badal, Y. T.; Sungkur, R. K. (2022) .Predictive modelling and analytics of students' grades using machine learning algorithms. Education and Information Technologies. https://doi.org/10.1007/s10639-022-11299-8

[2] N. Nithiyanandam et al., "Artificial Intelligence Assisted Student Learning and Performance Analysis using Instructor Evaluation Model," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 1555-1561, doi: 10.1109/ICESC54411.2022.9885462.

[3] M. Al-kmali, H. Mugahed, W. Boulila, M. Al-Sarem and A. Abuhamdah, "A Machine-Learning based Approach to Support Academic Decision-Making at Higher Educational Institutions," 2020 International Symposium on Networks, Computers and Communications (ISNCC), Montreal, QC, Canada, 2020, pp. 1-5, doi: 10.1109/ISNCC49221.2020.9297177.

[4] Anderson, T.; Anderson, R. (2017). Applications Of Machine Learning To Student Grade Prediction In Quantitative Business Courses. Global Journal of Business Pedagogy.Vol. 1.Number 3.

[5] H. Nakayama and A. Hattori, "Additional learning and forgetting by support vector machine and RBF networks," Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02., Singapore, 2002, pp. 1408-1412 vol.3, doi: 10.1109/ICONIP.2002.1202852.

[6] Iqbal, Z.; Qadir, J.; Mian, A. N.; and Kamiran, F. (2017). Machine Learning Based Student Grade Prediction: A Case Study. arXiv:1708.08744v1 [1010cs.CY] 17 Aug 2017

[7] Bujang, S. D. A.; Selamat, A.; Ibrahim, R.; Krejcar, O.; Herrera-Viedma, E.; Fujita, H.; & Ghani, N. A. Md. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access, 9, 95608–95621. https://doi.org/10.1109/access.2021.3093563

[8] M. D. Mubarak, Z. S. Ameen, A. S. Mubarak and F. Al-Turjman, "A Step Ahead Students CGPA Prediction Based on Support Vector Machines," 2022 International Conference on Artificial Intelligence in Everything (AIE), Lefkosa, Cyprus, 2022, pp. 189-192, doi: 10.1109/AIE57029.2022.00042.

[9] Yen-Jen Oyang, Shien-Ching Hwang, Yu-Yen Ou, Chien-Yu Chen and Zhi-Wei Chen, "A novel learning algorithm for data classification with radial basis function networks," Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02., Singapore, 2002, pp. 1021-1026 vol.2, doi: 10.1109/ICONIP.2002.1198215.

[10] Venkat, Naveen; & Srivastava, Sahaj; & Garg, Lakshya. (2018). Predicting Student Grades using Machine Learning. 10.13140/RG.2.2.21516.77449/1.

[11] Y. Nieto, V. Gacía-Díaz, C. Montenegro, C. C. González and R. González Crespo, "Usage of Machine Learning for Strategic Decision Making at Higher Educational Institutions," in IEEE Access, vol. 7, pp. 75007-75017, 2019, doi: 10.1109/ACCESS.2019.2919343.

[12] Polyzou, Agoritsa & Karypis, George. (2016). Grade Prediction with Course and Student Specific Models. 89-101. 10.1007/978-3-319-31753-3_8.

[13] Dhilipan, J.; Vijayalakshmi, N.; Suriya, S.; Christopher, Arockiya (2021). Prediction of Students Performance using Machine Learning. IOP Conf. Ser.: Mater. Sci. Eng. 1055 012122. doi:10.1088/1757-899X/1055/1/012122

[14] Prakash, Mr.; Garg, Sachin. (2021). Performance Analysis and Prediction Of Student Result Using Machine Learning. International Research Journal of Modernization in Engineering Technology and Science.Vol. 3.Issue 12.

[15] Z. Iqbal, A. Qayyum, S. Latif and J. Qadir, "Early Student Grade Prediction: An Empirical Study," 2019 2nd International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 2019, pp. 1-7, doi: 10.23919/ICACS.2019.8689136.

[16] Gadhavi, Mahesh; Patel, Chirag. (2017). Student Final Grade Prediction Based On Linear Regression. Indian Journal of Computer Science and Engineering (IJCSE). ISSN: 0976-5166. Vol. 8.No. 3.

[17] Yagci, M. (2022) Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learn. Environ. 9, 11. https://doi.org/10.1186/s40561-022-00192-z

[18] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[19] Microsoft Corporation. (2018). Microsoft Excel. Retrieved from https://office.microsoft.com/excel

[20] Dunlap, William & Jones, Marshall & Bittner, Alvah. (1983). Average correlations vs. correlated averages. Bulletin of the Psychonomic Society. 21. 213-216. 10.3758/BF03334690.

[21] Jernelv, Ine & Hjelme, Dag & Matsuura, Yuji & Aksnes, Astrid. (2020). Convolutional neural networks for classification and regression analysis of one-dimensional spectral data.

[22] Zeng, Kun & Liu, Jinlan & Jiang, Zhixia & Xu, Dongpo. (2022). A Scaling Transition Method from SGDM to SGD with 2ExpLR Strategy. Applied Sciences. 12. 12023. 10.3390/app122312023.

[23] Chai, Tianfeng & Draxler, R.R.. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature. Geoscientific Model Development. 7. 1247-1250. 10.5194/gmd-7-1247-2014.

[24] Gaftandzhieva, S.; Talukder, A.; Gohain, N.; Hussain, S.; Theodorou, P.; Salal, Y. K.; and Doneva, R. (2022) Exploring Online Activities to Predict the Final Grade of Student. Mathematics 2022, 10, 3758.https://doi.org/10.3390/math10203758