Received: 30 October 2024 | Revised: 19 December 2024 | Accepted: 22 January 2025 | Published online: 18 February 2025

RESEARCH ARTICLE

FinTech and Sustainable Innovation (FSI)
2025, Vol. 00(00) 1–12
DOI: 10.47852/bonviewFSI52024688

BON VIEW PUBLISHING

A Machine Learning and Digital Twin-Based Assistance System for Computer Science Students' Career Prediction and Activity Recognition

Tabassum Ferdous[1] and Mahfuzulhoq Chowdhury[1,*]

[1]*CSE Department, Chittagong University of Engineering and Technology, Bangladesh*

Abstract: The advancement of digital twin (DT), machine learning (ML), and deep learning (DL) technology has created new opportunities for students' activity recognition and career guidance in the educational sector. DT and ML technology can be used to update and monitor students' performance in various courses as well as daily life activity data. Existing research on student activity recognition and career suggestions did not create a dataset that included static, location, accelerometer, and academic data. They did not use DT technology for student performance monitoring. The existing ML-based works should have investigated various academic course data and job descriptions for the student's career recommendation with high accuracy. To deal with these issues, this paper creates an ML, DL, and DT-based assistance system for career recommendation and activity recognition based on student personal activity data and academic course results. Several ML and DL models were tested for career suggestion prediction, and logistic regression achieved the highest accuracy of 98%. The results exhibit that the convolutional neural network (CNN)-based DL model achieves the highest 96% accuracy for the student's daily activity recognition system. According to the performance comparison results, the proposed logistic regression-based career suggestion prediction system outperforms the existing works by at least 4% accuracy and 7% precision. The results also show that the proposed CNN-based activity recognition system achieves at least 5% higher accuracy and 3% higher precision values than previous works.

Keywords: machine learning, digital twin, career suggestion, activity recognition, deep learning

1. Introduction

The advancement of digital twin (DT) technology has given educational institutions and students a new perspective on career guidance (e.g., [1, 2]) student activity monitoring (e.g., [3, 4] and personalized learning (e.g., [5, 6]). Personalized learning entails tailoring education to each learner's unique needs and interests. A DT indicates the virtual representation of a physical object or entity that reproduces its characteristics and behaviors in a digital environment. In Rivera et al. [7], the authors discussed the prominent application of DTs in healthcare to improve interactions between systems, caregivers, and patients, as well as to continuously monitor patients' health using data-driven methods. In Tao et al. [8], the authors discussed the key components, development, research issues, and use cases of DT technology in a variety of industries (e.g., manufacturing, robotics, education, and healthcare). They also identified a number of challenges associated with DT-based system development, including data updates, security, and the coordination of physical and digital devices. In Subramanian et al. [9], the authors created a DT-based emotion classification system by combining several machine learning (ML) classifiers. In Gomerova et al. [10], the authors discussed how DT technology could be utilized to monitor student performance (e.g., attendance recording, course result updates, and face detection) in schools and universities. They also mentioned that technologies like DTs, ML, deep learning (DL) learning, and internet of things (IoT) can be combined to create an automated system for monitoring student performance and activity.

Continuous monitoring of academic performance can help identify areas where students may require additional assistance, resulting in improved learning outcomes. A career suggestion system is important for a computer science undergraduate student's future job life because it considers their skills and job descriptions (e.g., [11–13]). Because of their lack of experience, students frequently struggle to choose a career path after graduation.

Academic course results analysis, combined with skill analysis, is one of the most effective ways to predict a career path for computer science students. It should be noted that course teachers can test students' different programming and theoretical skills during various course examinations. To provide real-time assistance to computer science students, not only career advice but also the development of a physical activity monitoring system using DT technology is critical for personalized learning and growth. The integration of academic and real-time activity data provides a com-

***Corresponding author:** Mahfuzulhoq Chowdhury, CSE Department, Chittagong University of Engineering and Technology, Bangladesh. Email: mahfuz@cuet.ac.bd

prehensive view of the student, allowing for more accurate and relevant career advice. Recognition of students' daily activities can play a significant role in student performance improvement, class attendance tracking, real-time student status updates, and security, well-being, and health issues [6]. An automated system can collect various types of information, such as location and movement data, by using GPS and accelerometer data from students' mobile devices. Advanced learning models, such as ML and DL, can be used in a DT-based system to achieve high accuracy in classifying students' daily activities. In Gomes et al. [14], the authors created a K nearest neighbors (KNN) algorithm-based patient activity recognition system that achieved 79% accuracy. In Javeed et al. [15] and Parida et al. [16], the authors used DL and computer vision techniques, respectively, for human activity recognition. This work by Wan et al. [4] created a daily activity detection system using convolutional neural network (CNN) technology and smartphone sensors. Career guidance is an important part of a student's educational journey. The DT system can use ML algorithms to analyze academic performance and interests, resulting in personalized career recommendations. Because academic records are confidential, synthetic data can be used to simulate student performance. Several works (e.g., [1, 2, 5, 11–13]) are currently available in the field of ML-based student performance monitoring and career guidance. However, they did not create a real-time dataset containing computer science students' academic course results and personal information. They suffered from lower accuracy in career suggestions for students due to a lack of appropriate data preprocessing and hyperparameter tuning techniques. The previous student's performance monitoring system did not incorporate career recommendation and activity recognition using DT, ML, and DL technologies at the same time.

To surpass the previous challenges, this paper proposes a comprehensive DT system that combines static and dynamic data to provide personalized student monitoring. To provide accurate career recommendations based on academic performance as well as accurate activity recognition results, the proposed system implements and evaluates various ML, DT, and DL techniques. All ML and DL models were assessed by using accuracy value F1-score value, precision value, and some error metrics (such as MAPE, MASE, and RMSE). This article offers a comparison among proposed learning schemes to literary works for career suggestion and activity recognition.

The discussion of literary works is presented in Section two. Section three illustrates the proposed career suggestion and activity recognition model based on ML, DL, and DT technology. Section four presents the evaluation results of our proposed scheme, while section five discusses its conclusion.

2. Literature Review

This section highlights important literary works about career prediction and activity recognition. The work in Kumbhar et al. [1] created a smart career guidance system that employed DL techniques. Their system provided career guidance to degree students who may be unsure about their career paths. They gathered a dataset by administering a questionnaire that included information such as the user's skills, interests, and preferences. They did not improve students' academic performance. In Vignesh et al. [2], the authors used ML to predict a suitable department for students based on their performance and abilities. In their work, the KNN model achieves a higher accuracy of 94%, which is insufficient for this system. To assess the difficulty status of a lab course for university students, Hussain et al. [11] utilized SVM classifier. The limitation of their work is that they did not predict students' future job careers, and the model's accuracy is very low. The work by Sekeroglu et al. [12] predicted undergraduate students' performance using five ML classifiers. They also did not provide students with accurate career suggestions. According to the authors of Aboulsafa et al. [5] using real-time data and advanced learning processes can improve student learning experiences. They only created a DT system for the students. They did not, however, investigate the utilization of ML and DL learning models in career suggestion prediction and activity recognition systems. The work of Yağci et al. [13] investigated the performance of classical ML algorithms in predicting students' final grades in an undergraduate course. Their work is limited to Turkish students. Their proposed machine-learning model has a classification accuracy of less than 80%. The article in Ashalakshmi et al. [17] used CNN technology to make caregiver recommendations for 12th-grade students based on their skills and educational backgrounds. However, their method has an accuracy of only 81%. They did not work on CSE graduates' career recommendations due to a lack of job descriptions. The article in Hoti et al. [18] discussed some necessary factors associated with career recommendation for general students.

The study by Ouhaddou et al. [19] identified some factors associated with general students' future career path recommendations, including previous semester grades, demographics, social and economic status, learning factors, skills, and others. The work in Kadu et al. [20] created a random forest-based company prediction scheme for students that achieved an accuracy of 80%. The article by Goyal et al. [21] created a Chabot system for a student that includes job assistance and emotional support features. The authors of Yadav et al. [22] used an ANN classifier to predict career paths and achieved an accuracy of 95%. The study in Singh et al. [23] found that students' reading, writing, and mathematics subject scores are important for future career recommendations. The work by Soni et al. [24] used the KNN algorithm to analyze graduate students' resumes. They showed no accurate results for career prediction. The article in Jiang et al. [25] discussed the skills required for IT professionals to be satisfied with their jobs. However, they did not use any ML algorithms. Now we will go over some research on student activity recognition. The study by Gomes et al. [14] looked into collecting activity and intensity data from accelerometers to improve daily activity descriptions. They used the KNN algorithm to classify human activity and intensity levels. They did not investigate deep learning techniques for activity recognition. The works by Khan et al. [6] and Varshney et al. [3] used bagged trees and Long Short-Term Memory neural network (LSTM) models, respectively, for people's different types of activity recognition. The authors of Wan et al. [4] developed a DL-based system that uses the phone's accelerometer to track people's daily movements. The accuracy of their DL-based human activity tracking task is 94.8%. The work in Ghannem et al. [26] used fuzzy context analysis to recognize activities for elders. The work in Ullmann et al. [27] involved developing a radar-based activity recognition system using DL techniques. Their model has a lower accuracy (95%) and F1 score (92%) than other models in this domain. An improved LSTM-based human activity recognition model is shown in Li et al. [28]. Their model has accuracy and precision values of 91.65 and 91.75, respectively. The works in Ding et al. [29] and Thillaiarasu et al. [30] used WiFi signal and video for user activity recognition, respectively. The existing works did not create a DT that combined career guidance and human activity detection features. As previously discussed, existing research did not investigate both computer science undergraduate students' career suggestions and activity recognition using ML and DL techniques. To suppress the previous issues, this paper presents a DT-based assistance system

that uses ML and DL techniques to provide career suggestions and activity recognition to computer science undergraduate students.

3. Proposed Framework

Figure 1 delivers a general overview of our proposed DT-based student career suggestion and activity recognition system. By combining accelerometer data for activity recognition with academic data for career recommendations, the system provides a comprehensive and personalized user experience. Figure 2 exhibits a system diagram for our ML-based career suggestion system. Figure 3 depicts the activity recognition system in our proposed model. Both the career suggestion and activity recognition systems have several steps, including data collection, preprocessing, training and testing, model selection, and career suggestion and activity recognition using the chosen model. In the following, we have thoroughly discussed both ML-based career suggestion and DL-based activity recognition systems.

3.1. Dataset collection and preparation

The first step in our career suggestion and activity recognition prediction scheme is to collect data and prepare it. This career suggestion work dataset consists of 35,000 rows (each representing a student data) and 82 columns (academic subjects, course, or certificate information), capturing a diverse set of academic and interest-based metrics required for the analysis. The dataset's column contains the student's ID, grades for various academic courses and skill training courses, as well as labeled career information.

Figure 1
Methodology diagram of our digital twin-based assistance system

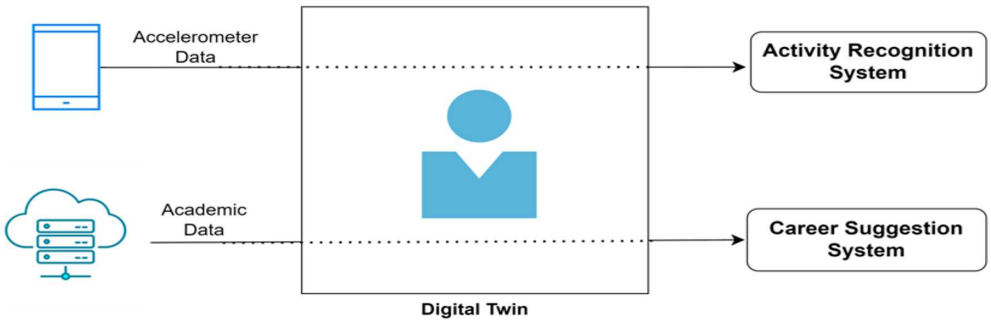


Figure 2
Overview of machine learning-based career suggestion prediction system

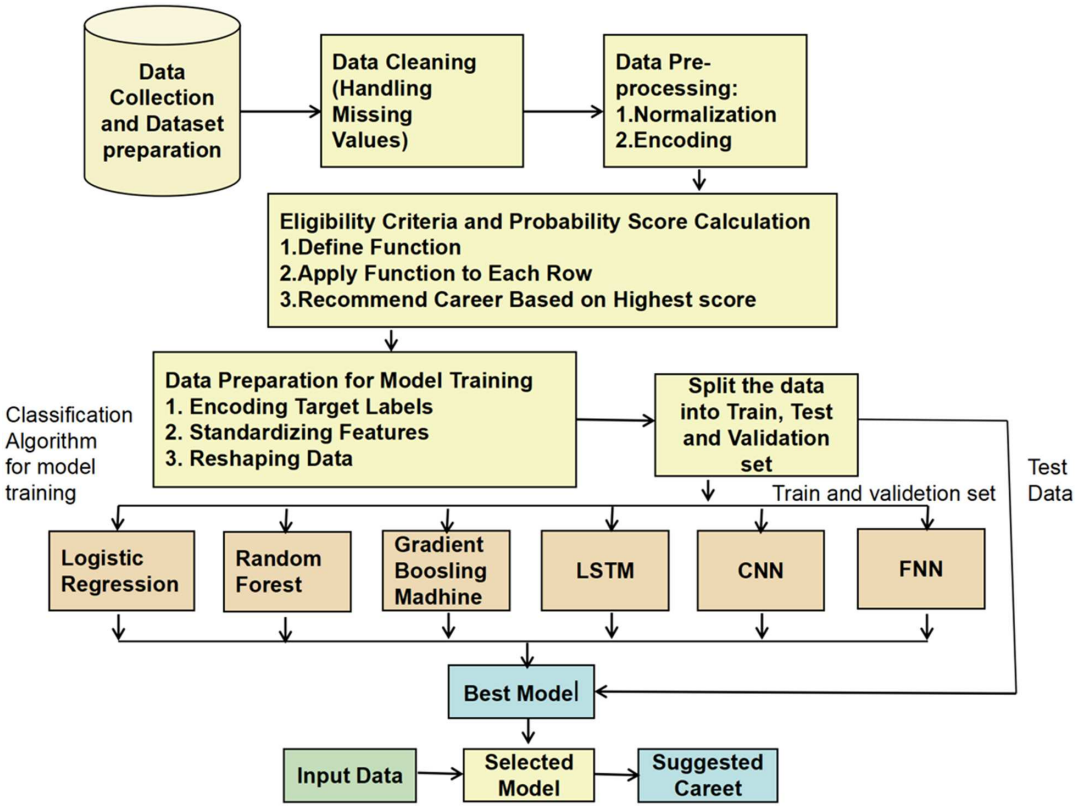


Figure 3
Overview of activity recognition system

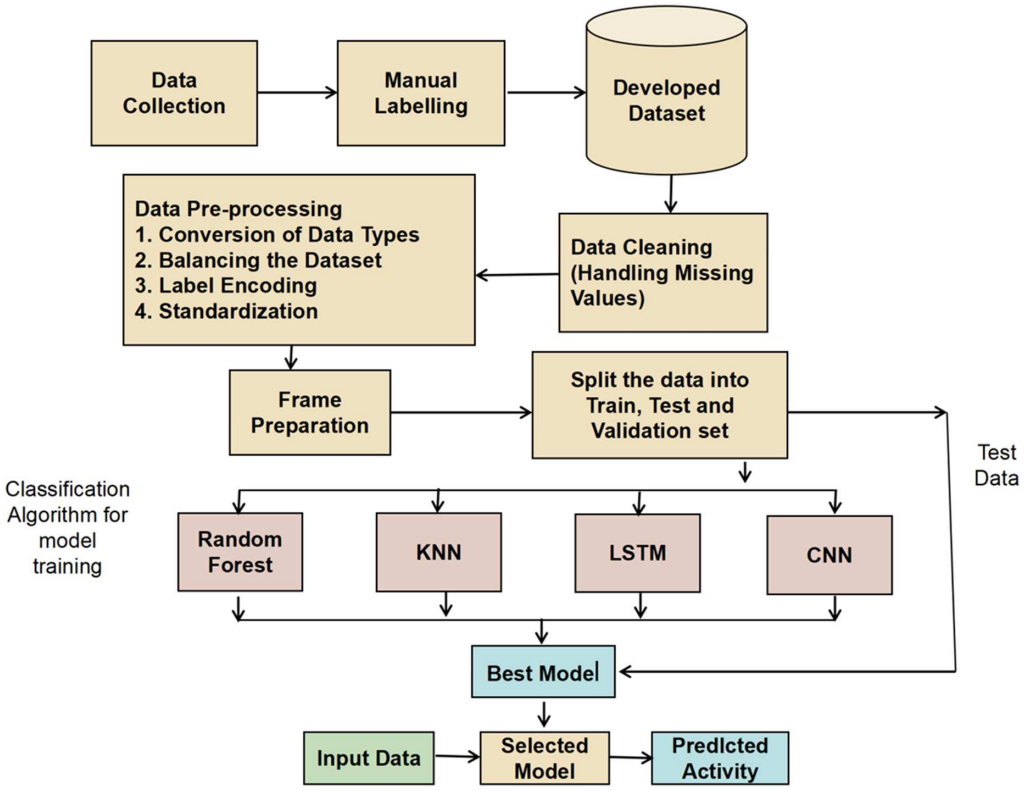
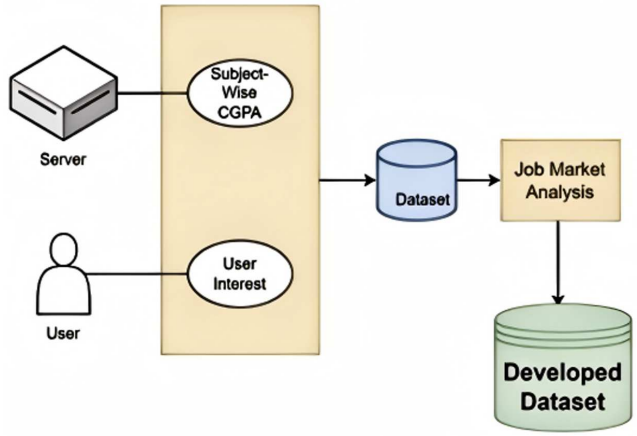


Figure 4 depicts the dataset development process for the career suggestion work. We gathered job information and requirements from a variety of online job posting sources. We identified 16 common careers for computer science graduate students (based on data from online newspapers and other sources). We also identified courses that are relevant to their career (see Figure 5). Then we collected grades from students for each related subject. Figure 6 shows a glimpse of our career suggestion dataset. Figure 7 highlights the activity recognition dataset. Figure 6 shows that grades for various subjects (courses and skill training certifications) range from 0 (lowest) to 4 (highest). The activity recognition data was manually entered into the "Sensor Logger" app on a Smartphone. The app was configured to record linear accelerometer data at a 100 Hz sampling rate (0.01 second per sample). The dataset contains 105,492 rows and six columns. The dataset includes information on four different activities: walking, downstairs, running, and upstairs. The column of the activity recognition dataset contains the timestamp of the recorded data, the time elapsed during recording, acceleration in the X, Y, and Z axis, and the labeled activity. The activity recognition dataset includes four distinct types of activity. The walking, downstairs, running, and upstairs data samples are 44634, 25244, 17962, and 17652, respectively.

Figure 4
Dataset development for career suggestion



3.2. Data preprocessing, labeling, and learning model apply

We cleaned the data, encoded the labels, and standardized the data as part of the preprocessing. In the data cleaning step, we checked for missing or null values and removed unnecessary columns from the dataset. Grades and certification values were normalized to a common scale to facilitate comparison and analysis. Categorical data (such as certifications) were encoded into numerical formats to facilitate ML model training and ensure compatibility with the algorithms chosen. Before training our models, we calculated a probability score for each career path. Specific eligibility criteria for each career path were established based on required academic performance in related subjects. For example, minimum grade levels were established for subjects relevant to each career path. A probability score for each career path was calculated using the student's performance in relevant subjects and activities. This entailed averaging grades and including additional scores for certifications and interests. We determined eligibility for a career path

Figure 5
Example of computer science student's course

1.Information Security Analyst Informaton Security Computer Networks Computer Networks (Sessional) Data Communication Data Commnication (Sessional) Applied Statistics & Queuing Theory Data Base Management Systems Data Base Management Systems (Sessional) Operating Systems	**3. Software Quality Engineer** Software Engineering Software Engineering (Sessional) Software Development Project (Sessional) Algorithms Design and Analysis Algorithms Design and Analysis (Sessional) Applied Statistics & Queuing Theory	**5. Software Architect** Software Architecture Software Engineering Sofware Engineering (Sessional) Discrete Mathematics System Analysis and Design System Analysis and Design (Sessional) Software Developments Project(Sessional)	**7.Software Engineer** Software Engineering Software Engineering (Sessional) Algorithms Design and Analysis (Sessional) Discrete Mathematics Data Structure Data Structure (Sessional) Object Oriented Programming Object Oriented Programming (Sessional)
2. Web Developer Internet Programming (Sessional) Object Oriented Programming Object Oriented Programming (Sessional) Operating Systems Operating Systems (Sessional) Data Base Management Systems Data Base Management Systems (Sessional)	**4.Database Administrator** Data Base Management Systems Data Base Manaement Systems (Sessional) Data Structure Data Structure (Sessional) Computer Architecture	**6. Date Sclentist** Applied Statistics & Queuing Theory Machine Learning Machine Learning (Sessional) Numerical Anahysis Numerical Analysis (Sessional) Artficial intelligence Artficial Intelligence (Sessional)	**8.Game Developer** Computer Graphics Computer Graphics (Sessional) Object Oriented Programming Object Oriented Programming (Sessional) Operating Systems Operating Systems (Sessional) Internet Programming (Sessional) Software Development wich JAVA (Sessional)

Figure 6
Glimpse of career suggestion dataset

	A	B	C	D	E	F	G	H	I
1	id	Structured Programm	Basic Electrical Eng	Differential Calculus	Physics	English	Computer Fundamer	Structured Programm	Basic Electrical Eng
2	1	2.5	2	3	3.5	3.25	3.75	2	3.25
3	2	4	3.75	3.75	2.25	2.75	3.75	3.25	3
4	3	3.75	3.75	2	3.5	2.75	3.75	3.5	2.5
5	4	3.75	2.5	2.5	2	2.25	3.5	3	3
6	5	4	3	3	3.25	4	2.75	4	2
7	6	4	4	3.5	2	2.5	2.75	4	2.25
8	7	3.5	3.25	4	3.5	2.75	3	2.75	2.5
9	8	3.25	4	3	3.5	2.25	2.75	3	2
10	9	2.75	2.25	2.75	2.5	3	3.25	2.5	2
11	10	2.5	2.25	2.75	3	2.75	2.25	4	2.75
12	11	4	3.75	2.75	3	4	3.25	3.75	2.5
13	12	3.25	3.5	4	3.75	2	2.75	4	3
14	13	2	2	2.75	2	3.75	3.5	3.5	3.75
15	14	2.5	2.5	2.5	3.75	3.75	3.25	2.75	2.25
16	15	2.5	3.75	2.5	3	2	2	2.5	4
17	16	2.5	2	2.25	3	3.75	3.75	2	3.25
18	17	3.75	3.25	2.75	3.75	2	3.5	2	2.75
19	18	2	3.75	2.5	4	2.25	3.75	2.25	2.75
20	19	2.5	4	3	4	3.75	3.25	3.25	2

Figure 7
Glimpse of activity recognition dataset

	A	B	C	D	E	F
1	time	seconds_elapsed	x	y	z	activity
2	1.72E+18	0.155419678	1.749961853	-0.355687737	-0.419590235	Downstair
3	1.72E+18	0.165572266	1.274217606	-0.326965094	-0.065240473	Downstair
4	1.72E+18	0.175717285	0.315225601	-0.256603003	-0.927318752	Downstair
5	1.72E+18	0.185861816	-0.059876442	-1.189276218	-0.742878556	Downstair
6	1.72E+18	0.196006592	0.28843689	-1.348096967	-0.312147856	Downstair
7	1.72E+18	0.206151367	-0.211243629	-0.724481344	-0.129105702	Downstair
8	1.72E+18	0.216296143	-0.939746857	-0.217891335	-0.282216907	Downstair
9	1.72E+18	0.226446289	-1.206015587	-0.047008395	-0.285644561	Downstair
10	1.72E+18	0.236591553	-1.261126518	0.011906981	-0.329352528	Downstair
11	1.72E+18	0.246766357	-2.65385294	0.172678947	0.178740561	Downstair
12	1.72E+18	0.256914795	-1.219870567	0.348408341	0.410764515	Downstair
13	1.72E+18	0.267027588	-0.926955223	0.644221544	-0.270963848	Downstair
14	1.72E+18	0.277172363	0.015420914	0.640550733	-0.272758186	Downstair
15	1.72E+18	0.287323975	0.073339462	0.675229669	0.025021642	Downstair
16	1.72E+18	0.297469482	0.285083771	0.548080206	0.327087522	Downstair
17	1.72E+18	0.307617188	0.375306129	0.530428052	0.298471212	Downstair
18	1.72E+18	0.317762695	0.356760834	0.347971559	0.296589732	Downstair
19	1.72E+18	0.327911133	0.276076317	0.18140614	0.486399531	Downstair

by reviewing the student's respective subject grades. If the students' related course grades are equal to or greater than three, they are qualified for the position. The mean grade value of those subjects was then calculated and assigned as a probability score value for that career or job posting (see Figure 8).

After calculating the probability score, each student's recommended career in the dataset was assigned based on their highest score (see Figure 9). This approach ensured that each student received a recommendation tailored to their strengths and interests. The dataset was prepared for ML model training by converting career names to numerical labels with Label Encoder. We used Standard Scaler to ensure that all features were on the same scale. We used grid search during the hyperparameter tuning process.

We have the dataset formatted for both ML and DL models. To ensure that the models were properly trained and evaluated, the career suggestion dataset was categorized into a training set (60%) and a validation set (20%). The test set data is 20%. Several ML and DL models were created and trained to predict the most appropriate career path for each student. The models used were logistic regression, random forest, gradient boosting machine, LSTM, CNN, and feed-forward neural network (FNN). To ensure that the activity recognition data was suitable for ML tasks, several preprocessing steps were carried out, including data cleaning, data type conversion, missing value handling, and dataset balancing.

We checked for null values and removed irrelevant columns from the activity recognition dataset, such as seconds elapsed and time. The x, y, and z columns were converted to floating-point numbers to make numerical analysis easier. Given the imbalance in activity samples, the dataset was balanced by restricting each activity to the size of the smallest class (17,652 samples). This was done to ensure that the ML model was not biased toward activities with a larger sample size. The string labels in the activity column were converted to numerical values via label encoding. This step was critical for the model to properly interpret the activity labels during training. The x, y, and z values were standardized to bring all features to a similar scale. This procedure entailed transforming the data to have zero mean and a standard deviation value of 1, which aids in the performance of ML models by ensuring uniformity across features. Then we labeled four activities. The preprocessed data were divided into frames to prepare it for model training. Each frame contained 200 observations (2 seconds of data), with a 50% overlap between adjacent frames. This segmentation allowed us to capture the temporal dynamics of the activities. The activity labels were converted into numeric values using label encoder, making

Figure 8
Glimpse of source code generation for eligibility checking and probability score generation

```
[ ] def eligibility_and_probabilities(row):
        if (row['Information Security'] >= 3 and row['Data Base Management Systems'] >= 3 and row['Operating Systems'] >= 3):
            row['Pr[Information Security Analyst]'] = row[['Information Security', 'Computer Networks', 'Computer Networks (Sessional)',
                                                'Data Communication', 'Data Communication (Sessional)', 'Applied Statistics & Queuing Theory',
                                                'Data Base Management Systems', 'Data Base Management Systems (Sessional)',
                                                'Operating Systems', 'Operating Systems(Sessional)']].mean()

        if (row['Internet Programming (Sessional)'] >= 3 and row['Data Base Management Systems'] >= 3):
            row['Pr[Web Developer]'] = row[['Internet Programming (Sessional)', 'Object Oriented Programming',
                                    'Object Oriented Programming (Sessional)', 'Operating Systems', 'Operating Systems(Sessional)',
                                    'Data Base Management Systems', 'Data Base Management Systems (Sessional)']].mean() + row['Web Dev - Interested or Certified']

        if (row['Software Engineering'] >= 3 and row['Algorithms Design and Analysis'] >= 3):
            row['Pr[Software Quality Engineer]'] = row[['Software Engineering', 'Software Engineering (Sessional)',
                                            'Software Development Project (Sessional)', 'Algorithms Design and Analysis',
                                            'Algorithms Design and Analysis (Sessional)', 'Applied Statistics & Queuing Theory']].mean() + row['SQA - Interested

        if (row['Software Engineering'] >= 3 and row['Algorithms Design and Analysis'] >= 3 and row['Software Architecture'] >= 3):
            row['Pr[Software Engineer]'] = row[['Software Engineering', 'Software Engineering (Sessional)', 'Algorithms Design and Analysis',
                                            'Algorithms Design and Analysis (Sessional)', 'Discrete Mathematics', 'Data Structure', 'Data Structure (Sessional)
                                            'Object Oriented Programming', 'Object Oriented Programming (Sessional)', 'Software Architecture']].mean()
```

Figure 9
Source code for career recommendation with high probability score

```
[ ] # Determine recommended career based on highest probability
    df['Recommended Career'] = df[['Pr[Information Security Analyst]', 'Pr[Web Developer]', 'Pr[Software Quality Engineer]', 'Pr[Software Engineer]',
```

```
[ ] df.head()
```

	id	Structured Programming	Basic Electrical Engineering	Differential Calculus and Integral Calculus	Physics	English	Computer Fundamentals (Sessional)	Structured Programming (Sessional)	Pr[Artificial Intelligence Engineer]	Pr[Cloud Engineer]	Pr[Robotics Engineer]	Pr[Not Eligible]	Recommended Career
0	1.0	2.50	2.00	3.00	3.50	3.25	3.75	2.00	3.142857	2.928571	NaN	NaN	Pr[Software Quality Engineer]
1	2.0	4.00	3.75	3.75	2.25	2.75	3.75	3.25	2.928571	3.107143	2.953125	NaN	Pr[Web Developer]
2	3.0	3.75	3.75	2.00	3.50	2.75	3.75	3.50	2.857143	2.642857	3.828125	NaN	Pr[Robotics Engineer]
3	4.0	3.75	2.50	2.50	2.00	2.25	3.50	3.00	NaN	NaN	7.078125	NaN	Pr[Robotics Engineer]

Figure 10
Label encoding for activity recognition

```
[ ]  from sklearn.preprocessing import LabelEncoder

     label = LabelEncoder()
     balanced_data['label'] = label.fit_transform(balanced_data['activity'])
     balanced_data.head()
```

	x	y	z	activity	label
0	-2.193040	1.476146	-0.687171	Walking	3
1	-2.246332	1.598482	-0.365511	Walking	3
2	-1.932697	1.432257	-0.248476	Walking	3
3	-1.505270	1.219969	-0.245012	Walking	3
4	-0.753758	0.940216	-0.192668	Walking	3

```
[ ]  activity_labels = dict(zip(label.classes_, label.transform(label.classes_)))
     print(activity_labels)
```

```
{'Downstair': 0, 'Running': 1, 'Upstair': 2, 'Walking': 3}
```

Figure 11
Scaling process on activity recognition

```
     X = balanced_data[['x', 'y', 'z']]
     y = balanced_data['label']
```

```
[ ]  scaler = StandardScaler()
     X = scaler.fit_transform(X)

     scaled_X = pd.DataFrame(data = X, columns=['x', 'y', 'z'])
     scaled_X['label'] = y.values

     scaled_X
```

	x	y	z	label
0	-0.564849	0.449264	-0.299420	3
1	-0.576230	0.479477	-0.212270	3
2	-0.509251	0.438425	-0.180560	3
3	-0.417971	0.385998	-0.179622	3

them more useful in model training (for example, walking activity is classified as level 3).

Figures 10 and 11 depict the label encoding and scaling process source code for our activity recognition project, respectively. During the model development phase, different ML and DL models were tested to determine the best algorithm for activity recognition. We evaluated the KNN, random forest, LSTM, and CNN models to determine the best model for activity recognition. The activity recognition dataset is divided into three categories: training (70%), validation (15%), and test set data (15%). The models were trained on the training set, tested with the testing dataset, and hyperparameter tuning (using the grid search technique) was applied to the validation set to improve model performance.

3.3. Model selection for career suggestion and activity recognition

We investigated various ML and DL models for career recommendation and activity recognition work. For career recommendation, models such as logistic regression, random forest, gradient boosting, CNN, LSTM, and FNN were tested. The analyzed ML and DL models for activity recognition include random forest, KNN, LSTM, and CNN. To select the best prediction model, we evaluated each model's accuracy value, precision value, recall value, F1 score value, mean absolute percentage error (MAPE) value, mean absolute scaled error (MASE) value, and Root mean squared error (RMSE) error results. We chose the best predictive models for career suggestions and activity recognition based on higher accuracy, precision, recall, and F1 scores.

4. Results and Discussion

In this section, we will compare different ML and DL models for career suggestion and activity recognition systems. Figure 12 evaluates the accuracy value, precision value, recall value, and F1 score values of different learning methods for the career suggestion system. We compared the performance of the logistic regression model, CNN model, FNN model, random forest model, and gradient boosting for career suggestion prediction results. The figure shows that the logistic regression model performs better than other literary models in terms of accuracy value (98%), precision value (95%), recall value (93%), and F1 score value (94%). Higher precision and recall values indicate fewer false positives and negatives. The lowest F1 score in our proposed logistic regression model is 94%, indicating a proper balance between the precision value and recall value. Compared to all other models, the LSTM method delivers the lowest accuracy value and precision results. Figure 13 exhibits that the logistic regression model has the lowest MAPE value (mean absolute percentage error) and RMSE value (root mean squared error) among all methods tested. The logistic regression model has the lowest RMSE value (0.97), indicating a high prediction accuracy. On a relative scale, the logistic regression model performs exceptionally well, with a MAPE value of 3.96%. Figure 13 shows that the LSTM model performs worst in terms of MAPE and RMSE values. Thus,

Figure 12
Performance comparison for career suggestion system

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ML	**Logistic Regression**	**98**	**95**	93	**94**
	Random Forest	96	94	92	93
	Gradient Boosting Machine	97	94	93	93
DL	LSTM	93	83	80	81
	CNN	97	93	94	93
	FNN	95	89	90	88

Figure 13
Error result comparison for a career suggestion system

	Model	MAPE (%)	RMSE
ML	**Logistic Regression**	**3.96**	**0.97**
	Random Forest	8.55	1.73
	Gradient Boosting Machine	9.16	1.63
DL	LSTM	14.57	2.29
	CNN	10.25	1.53
	FNN	9.13	1.76

Figure 14
Performance comparison for the activity recognition system

	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ML	KNN	94	91	96	92
	Random Forest	91.77	94	88	89
DL	LSTM	90	90	89	89
	CNN	**96**	**95**	95	**95**

Figure 15
Performance comparison for activity recognition system

	Model	MAPE (%)	RMSE	MASE
ML	KNN	1.496	0.47	0.08
	Random Forest	7.5	0.48	0.3
DL	LSTM	4.72	0.64	0.15
	CNN	1.48	**0.46**	**0.07**

based on Figures 12 and 13, we can conclude that logistic regression is the best model overall, as it outperforms both classification and regression metrics. The logistic regression model excels in accuracy value, precision value, recall value, and F1 score value, as well as having the lowest loss, MAPE, and RMSE, making it the most reliable and effective model for the given dataset.

Figure 14 compares the performance of various ML and DL models for student activity recognition (e.g., KNN, random forest, LSTM, and CNN). Figure 14 exhibits that the CNN model suppresses all other methods tested (e.g., KNN, random forest, LSTM) with an accuracy of 96%. An F1 score of 95% in the CNN model indicates an excellent balance of precision and recall. CNN models also have the highest precision value of 95%. Figure 14 also shows that the KNN model performs second best with an accuracy value of 94% and F score value of 92%. Figure 15 depicts the RMSE, MAPE, and MASE error values for various ML and DL models. The figure clearly shows that the CNN model produces better error values than other models. With an RMSE value of 0.46, MAPE value of 1.48, and MASE value of 0.07, the CNN model performs well in accurate activity predictions. The KNN model achieves the second-best results in terms of RMSE (.47), MASE (.08), and MAPE (1.496). Figures 14 and 15 highlight that the CNN model is the best predictive model for activity recognition because it has lower error values and higher accuracy. We created a web application to visualize DT-based students' academic data and career suggestions. Figure 16 depicts the front-end page of our web application. Figure 16 depicts the general information of computer science students as well as their CGPAs by semester. Figure 16 highlights a computer science student's grade in various courses during a semester. Figure 17 depicts the suggested career for a computer science student using a DT-based web application, as well as his or her extracurricular activity and location information.

Figure 16
(a) Webpage to visualize digital twin-based academic course data and (b) Webpage to visualize digital twin-based academic course data

(a)

Student Digital Twin

Name: Tabassum Ferdous
Date of Birth:29-12-2000
Grade: Level 4, Term II

Academic Records

Level1, Term I	Level1, Term II	Level2, Term I	Level2, Term II	Level3, Term I	Level3, Term II	Level4, Term I	Level4, Term II
3.8	3.9	3.7	3.6	3.8	3.7	3.9	4

(b)

Subject	CGPA
Data structure	3.5
Numerical Analysis	4.0
Electronic devices and circuits	3.0
Vector, calculus, linear algebra, complex variables	3.75
Engineering economics	3.5
Data structure (sessional)	3.25
Numerical Analysis (Sessional)	2.75
Electronic devices and circuits (Sessional)	3.0
Engineering Drawing and CAD (Sessional)	4.0

Figure 17
Webpage to visualize suggested career based on academic course data

Suggested career	**Web developer**
Extracurricular activities	**Computer club (general member), greater mymensingh association (religious secretary)**
Behavior and attendance	**Attendance: 95 percent** **Participation: High**
Location Details	**Ladies hall road, CUET, Raouzan, Bangladesh**

Figure 18 depicts a comparison of results among the proposed logistic regression-based career suggestion system with other related works. We compared our findings to three existing studies [2, 11,13]. The comparison results show that our proposed logistic regression-based prediction model outperforms existing models by at least 4% in accuracy, 7% in precision, 1% in recall, and 5% in F1 score. The existing work S. Vignesh et al. [2] using the KNN model yields the second-best career prediction results in terms of accuracy value. The main reason for the proposed logistic regression scheme's superiority is that our model generates a real-time dataset by collecting grade results from students and assessing necessary skills for a computer science career. We used proper preprocessing, normalization, best model selection, and hyperparameter tuning techniques to achieve the best career suggestion prediction result. Figure 19 compares the proposed CNN model-based activity prediction system's results to those of previous works [15, 16]. The comparison results exhibit that the proposed CNN-based prediction model surpasses the existing works by at least 5% in accuracy, 3% in precision, 5% in recall, and 4% in F1 score. Because of proper dataset collection for student activity,

Figure 18
Comparison results career suggestion system with existing works

Author	Approach	Accuracy (%)	Precision	Recall	F1 score
M. Yagci et al. [13]	Random forest	75	75.2	74.6	74.9
M. Hussain et al. [11]	SVM	80	83	92	87
S. Vignesh et al., [2]	KNN	94	88	90	89
Proposed Model	**Logistic regression**	**98**	**95**	**93**	**94**

Figure 19
Comparison results activity recognition system with existing works

Author	Approach	Accuracy (%)	Precision	Recall	F1 score
M. Javeed et al. [15]	Deep learning	88.57	90	88	89
L. Parida et al. [16]	Conv. LSTM	90.93	92	90	91
Proposed Model	**CNN**	**96**	**95**	**95**	**95**

proper preprocessing, feature extraction, and hyperparameter tuning techniques, the CNN model outperforms the other compared works.

5. Conclusion

This paper describes a DT-based assistance system for undergraduate computer science students that uses ML technologies to provide features such as career suggestions and activity recognition. The proposed system creates an appropriate dataset for the student assistance system by effectively integrating accelerometer data, location data, and academic data to provide students with personalized career guidance. The proposed system investigated various ML and DL techniques for career recommendation and activity recognition. The career suggestion prediction results revealed that logistic regression techniques outperformed other ML and DL models by achieving 98% accuracy, 95% precision, 94% F1 score, and 0.97 RMSE values. The proposed CNN model outperformed other ML and DL models in activity recognition prediction, with 96% accuracy, 95% precision, and 0.46 RMSE values. The comparison results on the career suggestion task indicated that the proposed CNN model achieves at least 4% more accuracy, 7% more precision, and 5% more F1 score values than previous works. The comparison results on the activity recognition task show that the proposed CNN model outperforms the existing works by 5% accuracy, 3% precision, and 4% F1 score value. In the future, we hope to create a mobile application for students that includes features such as emergency help assistance through police contact, mental health assistance, and an Artificial intelligence (AI)-based Chabot for students' personal guidance, among others. The future extension of this work will include the development of a block chain-based secure data exchange and user authentication system for monitoring student results and career guidance. In the future, we hope to incorporate more activity recognition for students through DL and IoT techniques such as sleeping and studying activities, among others.

Recommendation

The findings hinted that logistic regression techniques outperformed are most suitable for career suggestion systems and CNN for students' activity recognition work.

Acknowledgement

The authors are grateful to CUET, CSE department for research facilities.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work is available upon reasonable request from the corresponding author.

Author Contribution Statement

Tabassum Ferdous: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources,

Visualization, Project administration. **Mahfuzulhoq Chowdhury:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration.

References

[1] Kumbhar, V. R., Maddel, M. M., & Raut, Y. (2023). Smart model for career guidance using hybrid deep learning technique. *2023 1st International Conference on Innovations in High Speed Communication and Signal Processing*, 327–331. https://doi.org/10.1109/IHCSP56702.2023.10127152

[2] Vignesh, S., Shivani Priyanka, C., Shree Manju, H., & Mythili, K. (2021). An intelligent career guidance system using machine learning. *2021 7th International Conference on Advanced Computing and Communication Systems*, 987–990. https://doi.org/10.1109/ICACCS51430.2021.9441978

[3] Varshney, N., Bakariya, B., Kushwaha, A. K. S., & Khare, M. (2022). Human activity recognition by combining external features with accelerometer sensor data using deep learning network model. *Multimedia Tools and Applications*, *81*(24), 34633–34652. https://doi.org/10.1007/s11042-021-11313-0

[4] Wan, S., Qi, Q., Xu, X., Tong, C., & Gu, Z. (2020). Deep learning models for realtime human activity recognition with smartphones. *Mobile Networks and Applications*, *25*, 743–755. https://doi.org/10.1007/s11036-019-01445-x

[5] Aboulsafa, E. I., Khayat, G. A. E., & Elmorsy, S. A. (2023). An educational human digital twin proposed model for personalized E-learning. *2023 IEEE Afro-Mediterranean Conference on Artificial Intelligence*, 1–8. https://doi.org/10.1109/AMCAI59331.2023.10431503

[6] Khan, R., Abbas, M., Anjum, R., Waheed, F., Ahmed, S., & Bangash, F. (2020). Evaluating machine learning techniques on human activity recognition using accelerometer data. *2020 International Conference on UK-China Emerging Technologies*, 1–6. https://doi.org/10.1109/UCET51115.2020.9205376

[7] Rivera, L. F., Jiménez, M. A., Angara, P. P., Villegas, N. M., Tamura, G., & Müller, H. A. (2019). Towards continuous monitoring in personalized healthcare through digital twins. *In 29th Annual International Conference on Computer Science and Software Engineering*, 329–335. https://dl.acm.org/doi/10.5555/3370272.3370310

[8] Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. *In IEEE Transactions on Industrial Informatics*, *15*(4), 2405–2415. https://doi.org/10.1109/TII.2018.2873186

[9] Subramanian, B., Kim, J., Maray, M., & Paul, A. (2022). Digital twin model: A real-time emotion recognition system for personalized healthcare. *In IEEE Access*, *10*, 81155–81165. https://doi.org/10.1109/ACCESS.2022.3193941

[10] Gomerova, A., Volkov, A., Muratchaev, S., Lukmanova, O., & Afonin, I. (2021). Digital twins for students: Approaches, advantages and novelty. *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering*, 1937–1940. https://doi.org/10.1109/ElConRus51938.2021.9396360

[11] Hussain, M., Zhu, W., & Zhang, W., et al. (2019). Using machine learning to predict student difficulties from learning session data. *Artificial Intelligence Review*, *52*, 381–407. https://doi.org/10.1007/s10462-018-9620-8

[12] Şekeroğlu, B., Dimililer, K., & Tuncal, K. (2019). Student performance prediction and classification using machine learning algorithms. *In 8th International Conference on Educational and Information Technology*, 7–11. https://dl.acm.org/doi/10.1145/3318396.3318419

[13] Yağci, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*, 11. https://doi.org/10.1186/s40561-022-00192-z

[14] Gomes, E., Bertini, L., Campos, W. R., Sobral, A. P., Mocaiber, I., & Copetti, A. (2021). Machine learning algorithms for activity-intensity recognition using accelerometer data. *Sensors*, *21*(4), 1214. https://doi.org/10.3390/s21041214

[15] Javeed, M., & Jalal, A. (2023). Deep activity recognition based on patterns discovery for healthcare monitoring. *2023 4th International Conference on Advancements in Computational Sciences*, 1–6. https://doi.org/10.1109/ICACS55311.2023.10089764

[16] Parida, L., Parida, B. R., Mishra, M. R., Jayasingh, S. K., Samal, T., & Ray, S. (2023). A novel approach for human activity recognition using vision based method. *2023 1st International Conference on Circuits, Power and Intelligent Systems*, 1–5. https://doi.org/10.1109/CCPIS59145.2023.10292055

[17] Ashalakshmi, R., & Hemalatha, S. (2023). Implementing convolutional neural networks for career prediction: A case study on twelfth grade students. *2023 International Conference on Emerging Research in Computational Science*, 1–6. https://doi.org/10.1109/ICERCS57948.2023.10433993

[18] Hoti, A., & Zenuni, X. (2024). Factors influencing student academic performance and career choices. *2024 8th International Artificial Intelligence and Data Processing Symposium*, 1–8. https://doi.org/10.1109/IDAP64064.2024.10710702

[19] Ouhaddou, C., Retbi, A., & Bennani, S. (2023). A framework for predicting student academic path using machine learning. *2023 7th IEEE Congress on Information Science and Technology*, 425–430. https://doi.org/10.1109/CiSt56084.2023.10409990

[20] Kadu, R., Assudani, P. J., Mukewar, T., Kapgate, J., & Bijekar, R. (2024). Student placement prediction and skill recommendation system using machine learning algorithms. *2024 International Conference on Inventive Computation Technologies*, 401–408. https://doi.org/10.1109/ICICT60155.2024.10544738

[21] Goyal, R., Chaudhary, N., & Singh, M. (2023). Machine learning based Intelligent Career Counselling Chatbot (ICCC). *2023 International Conference on Computer Communication and Informatics*, 1–8. https://doi.org/10.1109/ICCCI56745.2023.10128305

[22] Yadav, A. K., Dixit, A., Tripathi, A., Chowdhary, S. K., & Jangra, V. (2023). Career prediction system using ANN MLP classifier. *2023 14th International Conference on Computing Communication and Networking Technologies*, 1–7. https://doi.org/10.1109/ICCCNT56998.2023.10307057

[23] Singh, M., Pattanaik, S., Singh, G., Luthra, P., Singla, S., & Sharma, B. (2023). Student career prediction using machine learning. *2023 International Conference on Advanced Computing & Communication Technologies*, 473–477. https://doi.org/10.1109/ICACCTech61146.2023.00083

[24] Soni, N., Nishant, N., Singh, M., Ansari, S., Siddiqui, A. T., & Ansar, S. A. (2023). Leveraging machine learning for career prediction and resume analysis in career assist. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering*, 1009–1014. https://doi.org/10.1109/UPCON59197.2023.10434306

[25] Jiang, J., Huang, W. W., Klein, G., & Tsai, J. C.-A. (2020). The career satisfaction of IT professionals with mixed job demands. *In IEEE Transactions on Engineering Management*, *67*(1), 30–41. https://doi.org/10.1109/TEM.2018.2870085

[26] Ghannem, A., Francis, E., Nabli, H., Sliman, L., & Djemaa, R. B. (2023). Fuzzy FCA-based elderly activity recognition. *2023 International Conference on Computer and Applications*, 1–6. https://doi.org/10.1109/ICCA59364.2023.10401417

[27] Ullmann, I., Guendel, R. G., Kruse, N. C., Fioranelli, F., & Yarovoy, A. (2023). Radar-based continuous human activity recognition with multi-label classification. *2023 IEEE SENSORS*, 1–4. https://doi.org/10.1109/SENSORS56945.2023.10324957,

[28] Li, W., Sun, X., He, T., & Jiang, T. (2024). Development of a human activity recognition algorithm based on BiLSTM for construction workers. *2024 7th International Conference on Advanced Algorithms and Control Engineering*, 198–204. https://doi.org/10.1109/ICAACE61206.2024.10548765

[29] Ding, X., Mei, Y., Cai, B., Zhou, Y., Yu, J., Xie, W., & Jiang, T. (2024). A novel multimodal human activity recognition based on self-attention mechanism. In *2024 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 1–6. https://doi.org/10.1109/BMSB62888.2024.10608274

[30] Thillaiarasu, N., Satish, P., Bhargava, V., Reddy, S.V., & Radhakrishnan, A. (2024). Augmented home security system with computer vision based human activity recognition. *2024 10th International Conference on Advanced Computing and Communication Systems*, 2329–2333. https://doi.org/10.1109/ICACCS60874.2024.10716826

How to Cite: Ferdous, T., & Chowdhury, M. (2025). A Machine Learning and Digital Twin-Based Assistance System for Computer Science Students' Career Prediction and Activity Recognition. *FinTech and Sustainable Innovation (FSI)*. https://doi.org/10.47852/bonviewFSI52024688