IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE EDUCATION SOCIETY SECTION

Received 23 May 2023, accepted 10 June 2023, date of publication 22 June 2023, date of current version 28 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3288695

RESEARCH ARTICLE

Predicting Behavior Change in Students With Special Education Needs Using Multimodal Learning Analytics

ROSANNA YUEN-YAN CHAN[1,2], (Senior Member, IEEE), CHUN MAN VICTOR WONG[3], AND YEN NA YUM[3]

[1]Centre for Perceptual and Interactive Intelligence, The Chinese University of Hong Kong, Hong Kong, SAR
[2]Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, SAR
[3]Department of Special Education and Counseling, The Education University of Hong Kong, Hong Kong, SAR

Corresponding author: Rosanna Yuen-Yan Chan (yychan@ie.cuhk.edu.hk)

This work was supported in part by the Centre for Perceptual and Interactive Intelligence (CPII) Ltd., under the Innovation and Technology Commission (ITC)'s InnoHK.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Caritas Resurrection School under Reference No. 2020-2021-0010.

ABSTRACT The availability of educational data in novel ways and formats brings new opportunities to students with special education needs (SEN), whose behaviour and learning are highly sensitive to their body conditions and surrounding environments. Multimodal learning analytics (MMLA) captures learner and learning environment data in various modalities and analyses them to explain the underlying educational insights. In this work, we applied MMLA to predict SEN students' behaviour change upon their participation in applied behaviour analysis (ABA) therapies, where ABA therapy is an intervention in special education that aims at treating behavioural problems and fostering positive behaviour changes. Here we show that by inputting multimodal educational data, our machine learning models and deep neural network can predict SEN students' behaviour change with optimum performance of 98% accuracy and 97% precision. We also demonstrate how environmental, psychological, and motion sensor data can significantly improve the statistical performance of predictive models with only traditional educational data. Our work has been applied to the Integrated Intelligent Intervention Learning (3I Learning) System, enhancing intensive ABA therapies for over 500 SEN students in Hong Kong and Singapore since 2020.

INDEX TERMS Applied behavior analysis (ABA), multimodal learning analytics (MMLA), predictive modeling, special education needs (SEN).

I. INTRODUCTION

Students with special education needs (SEN) often exhibit behavioural characteristics such as hyperactivity, short attention span, and emotional liability. Many are also at risk for academic and social problems [1]. Research suggests that inappropriate behaviours in SEN students, such as those with autism spectrum disorders (ASD), are associated with abnormalities in brain development [2]. Besides, attention deficit hyperactivity disorder (ADHD) and some learning disabilities also have their genetic origin [3]. Contextually inappropriate behaviours (such as aggression and self-harm) can hinder SEN students' social and personal development. Therefore, promoting positive behaviours is an important learning outcome in special education.

Applied behaviour analysis (ABA) therapy is an intervention approach aiming at SEN students' behaviour change [4]. ABA strategies are designed based on behavioural science and principles such as reinforcement and stimulus control. Through promoting desirable behaviour change,

The associate editor coordinating the review of this manuscript and approving it for publication was John Mitchell.

socially significant outcomes can be facilitated [5]. Recently, Alves et al. offered a systematic review of ABA technologies [6], including support systems for ABA applications (p.118667). The reviewed works ranged from web-based services and data visualisation for teaching children with low-functioning autism [7] to real-time monitoring [8] and data management [9] for personalised intervention. However, a dearth of works targeting ABA outcomes prediction exists. It is worth noting that the behaviour analysis processes in ABA therapy are evidence-based and highly systematic. This nature makes data-driven techniques such as learning analytics (LA) suitable for enhancing ABA-related technologies. Meanwhile, LA is often employed in educational practice to understand and optimise learning and the learning environment [10], giving it the potential to enhance existing ABA practice.

This work aims to enhance existing ABA therapy by predicting SEN students' behaviour change using educational data in multiple modalities. In particular, our study is guided by the following research questions.

- **RQ1** What are the statistical characteristics of ambient environmental, physiological, and motion data collected from SEN students' ABA therapy sessions?
- **RQ2** Can sensors and wearable data enhance the prediction of SEN students' behaviour change over traditional educational data?
- **RQ3** Can machine learning (ML) algorithms be applied to MMLA for SEN students' behaviour change prediction, and what is their performance compared with other existing works in MMLA?

The above questions will be answered thoroughly in Section IV and Section V of the current paper. Our work's contributions include the following:

- We design and develop a multimodal data collection system for ABA therapies, collect and analyse data from 1,130 ABA therapy sessions, and provide detailed statistical interpretations of our results.
- We show, with statistical evidence, that sensors and wearable data can significantly enhance the prediction of SEN students' behaviour change over traditional educational data.
- We demonstrate that ML algorithms and deep neural networks (DNN) can predict SEN students' behaviour change accurately. We also provide extensive performance evaluations of our predictive models and benchmark our results with other existing works.

Our research will provide new insights into ABA practices, especially in predicting students' learning with the help of the Internet of Things (IoT) sensors and wearables. Through this work, the broad engineering community will further realise the application of MMLA to enhance behavioural interventions in SEN students and promote their skills acquisition. The new findings presented in this article also provide valuable references for future research in technologies for special education.

II. THEORETICAL BACKGROUND

A. APPLIED BEHAVIOR ANALYSIS

Applied Behavior Analysis (ABA) is an intervention method in which pedagogical strategies derived from the principles of behaviour are systematically applied to promote socially significant behaviours and reduce problem behaviours [4]. The set of basic principles, which are statements about how environmental variables act as input to a function of behaviour, have been evaluated scientifically by experimental analyses of behaviours (p.155). In ABA, behaviour is viewed as the learner's interaction with his or her surrounding environment and involves the movement of some part(s) of the learner's body. Learning behaviour occurs within the environmental context. At the same time, the learning environment is regarded as the full set of physical circumstances in which the learner is situated.

The learning outcome of ABA lessons is the achievement of behaviour changes that improve learners' quality of life in communication and daily living skills. A systematic and measurable behaviour assessment scheme is defined before the ABA lessons. The target behaviour is often broken down into smaller tasks, while positive reinforcements are often used to encourage goal achievement. Assessment criteria include whether the target task is achieved (plus) or not (minus), whether a prompt from the therapist (prompt) is needed to facilitate task achievement, or if the student is behaving in a way that is unrelated to the task (off task). Furthermore, behaviour change is effective if it is durable over time [11]. Therefore, a subsequent follow-up reassessment of the developed behaviour is needed to ensure the effectiveness of the therapy.

B. FACTORS AFFECTING SEN STUDENTS' LEARNING

1) AMBIENT ENVIRONMENTAL FACTORS

Students with special needs can be susceptible to ambient environmental conditions due to their dysfunction in sensory processing. A previous study showed that high levels of CO2 content caused fatigue and difficulties in concentration in SEN students, especially those with ADHD [12]. Another study performed with intellectually disabled preschool students revealed that classroom thermal discomfort (e.g., high nearby ambient temperature) could distract them from learning and influence their mood and health [13]. The same study also suggested that students with intellectual disabilities (ID) are more vulnerable to acoustic discomforts due to their psychologically stressful conditions (p.115). Researchers also studied the relationship between classroom lighting and SEN students' comfort. They found that inappropriate lighting and glare affect individual SEN students to different extents, while they felt tired and irritated because of lighting discomfort, in general [14]. However, teachers and therapists often have no control over lighting characteristics except switching on or off (p.105).

2) PHYSIOLOGICAL FACTORS

Emotion can affect learning and engagement in students with and without SEN. In particular, students with ID often exhibit

anxiety due to internal stress. Blood pressure, body temperature, and heart rate are physiological markers for stress that hinder learning [15]. It was shown that mild conditions could reduce these inhibitors in SEN students [16]. It is known that abnormally high or low levels of skin conductance (measured through galvanic skin response, GSR) hindered the learning performance of SEN students [17]. Besides, a study also found that body movement facilitated by motion-based technology positively impacted SEN students' short-term memory skills [18].

C. MULTIMODAL LEARNING ANALYTICS

MMLA employs multiple sources and formats of educational data such as activity logs, audio, video and biosensors to enrich learning analytics [19]. MMLA is significantly enhanced by the Internet of Things (IoT) technologies because the latter allows convenient capturing of multimodal data from the complex learning environment [20]. Multimodal educational data collected by IoT sensors include those detecting learners' motion (e.g., head and body) and physiological (e,g., heart, brain, and skin) behaviour, as well as those measuring the ambient learning environment (e.g., light, humidity, temperature, and noise). These data were collected from physical objects or human bodies, then encoded into a machine-interpretable format and served as input to MMLA [21]. Possible interpretations of the observed learning process can be assigned based on validated learning theories.

MMLA has been used to study a variety of learning goals for SEN students. For example, motion sensor data were combined with cognitive skills measurement data (short-term memory, visual processing, and crystallised knowledge) to evaluate the effect of movement-based educational games on SEN students' academic performance [22]. Body movement log data were also compared with teachers' observations and interviews in physically impaired students' psychomotor abilities and psychomotor speed gains [23]. Besides, multimodal (audio, video, and autonomic physiology) learning data collected during robot-assisted therapy were evaluated to estimate the effect and engagement of children with autism [24]. In a recent study, wearable biosensors were employed to collect peripheral physiological (cardiovascular and electrodermal activity) and motion (accelerometer) signals of youth with ASD to predict their aggression behaviours [25].

A few existing works combined MMLA and machine learning (ML) to develop predictive models for various learning outcomes. For example, using multimodal data obtained from natural language processing (NLP) and video recognition to detect students' impasses during collaborative problem-solving [26]; using data such as seat pressure, heart rate, and facial expression to detect students' drowsiness during e-learning lessons [27], and predicting computer science students' course performance by motion data in addition to traditional demographics and educational data [28]. Other related works [29], [30], [31], [32] are listed in Table 15.

TABLE 1. Baseline demographics and diagnosed conditions of the participants.

Variable	n	%
Sex		
Female	6	17%
Male	29	83%
Age		
1 to 5 years	10	28.57%
6 to 10 years	16	45.71%
11 to 15 years	9	25.71%
School type		
Therapy centre	14	40%
Special school	21	60%
Diagnosed SEN condition(s)†		
Mild Autism Spectrum Disorders	15	42.86%
Moderate Autism Spectrum Disorders	18	51.43%
Mild Intellectual Disability	6	17.14%
Moderate Intellectual Disability	15	42.86%
Speech Delay	2	5.71%
General Developmental Delay	2	5.71%

†A participant may be diagnosed with more than one condition.

III. METHOD

A. CONTEXT

The current study was conducted in ABA therapy sessions carried out between students and teachers in a one-to-one manner. Each session was targeted at a behavioural task in one of the following six domains:

1) Academic and Learning
2) Communication
3) Social Emotion
4) Sensory Motor Skills
5) Independent and Self-help
6) Behavioural Development

Each task was further broken down into a chain of behaviour components that the student could have already performed with a little support. The training sessions were around 20 - 30 minutes long. Following the ABA practice, our sessions comprised behaviour-analytic-based instruction procedures consisting of antecedent (A) stimulus, behaviour (B), and consequences (C) events. An antecedent is a stimulus in the student's environment before a target behaviour occurs. A behaviour is an activity that the student does. A consequence is a stimulus following the behaviour that changes immediately according to the behaviour.

B. PARTICIPANTS AND PROCEDURE

The participants were thirty-two SEN students from two K12 special schools and one preschool education centre. Their baseline characteristics are given in Table 1. The written consent of every participant's parent or guardian was obtained prior to the commencement of the study. The steps below were performed between a participant and a therapist:

1) The system recommends a target behavioural response.
2) The therapist teaches the recommended task by:
   a) presenting one or more stimuli; and
   b) observing the student's response.

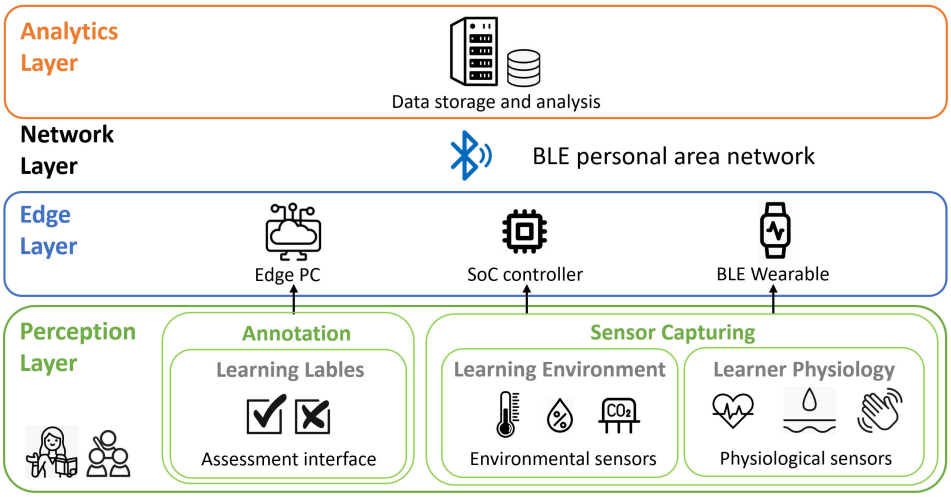

FIGURE 1. Overall 4-tier IoT system architecture for data collection.

3) The therapist observes the student's response. Whenever necessary, the therapist provides a prompt (such as a verbal or gestural message or physical guidance) to facilitate the student's correct response.
4) The therapist observes and assesses the student's response.
5) Steps 2 to 4 are repeated recursively until the student has mastered the target behaviour or the session ends.

In addition, maintenance probe sessions were conducted six months after the training of individual tasks to determine whether the mastered behavioural skills could be maintained over time. In the end, a total number of 1,130 within-subject therapy-probe session pairs were obtained.

C. SYSTEM DESIGN AND DEVELOPMENT

We have developed an IoT-based system to collect multimodal educational data arising from ABA therapy sessions. The overall system architecture is illustrated in Fig. 1. Our system has been integrated into the Integrated Intelligent Intervention-learning system (3I Learning system) [34] developed by the authors of the current work. The system consists of four tiers, namely the perception layer, edge layer, network layer, and analytics layer:

- The *perception layer* includes physiological sensors (that detect the participants' physiological conditions, including skin temperature, heart beat rate, skin conductance, and motion in terms of acceleration in three dimensions) and environmental sensors (that sense the indoor carbon dioxide concentration, light intensity, temperature, and humidity) (Fig. 2). Besides, this layer also includes a client interface that allows the therapists to enter their assessment results of the learner's behaviour responses (Fig. 3).
- The *edge layer* includes a Bluetooth low energy [BLE] wearable (Empatica E4 wristband), a system on a chip (SoC) controller, and an edge tablet personal computer (edge PC). The wearable contains physiological sensors. The SoC controller contains environmental sensors, and the edge PC provides an assessment interface for the therapists to input their assessment results.
- The *network layer* consists of a BLE personal area network connecting the edge layer devices (the edge PC, SoC controller, and the Empathica E4 wristband) to the analytics layer.
- The *analytics layer* contains a tablet PC that temporarily stores the gathered data and transmits it to the cloud server. It also provides a simple visualisation of the measurement values.

FIGURE 2. Perception layer equipment set of the current study. It includes an Empatica E4 wristband (left), a IoT sensors box with SoC controller (right), an edge tablet PC for assessment labels input (bottom), and a display screen of real-time data (top).

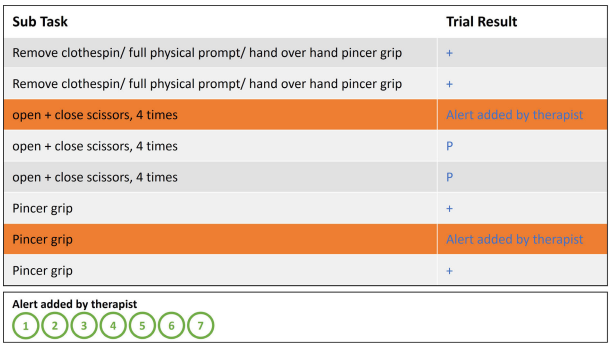

FIGURE 3. Assessment interface of the 3I Learning system installed on an edge tablet PC in the perception layer.

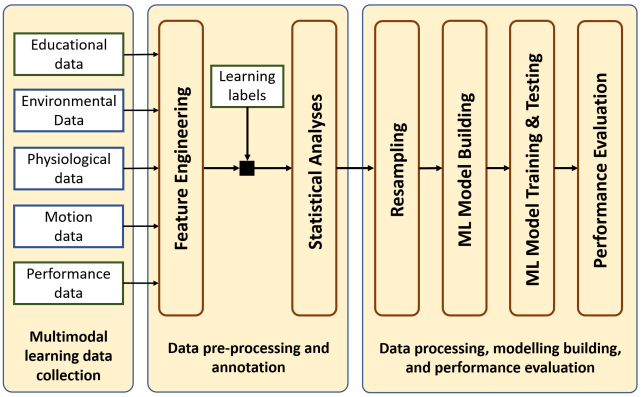

FIGURE 4. The overall workflow of the current study.

IV. THE CURRENT WORK

We frame our prediction problem as a binary classification task. The inputs are educational data, ambient environmental data, physiological data, and motion data. The prediction target is behaviour change achievement, represented by the *Changed* class and the *No Changed* class. The overall workflow for our MMLA is summarised in Fig. 4. It consists of three main stages, namely:

1) *Multimodal learning data collection:* This includes the performance of the ABA therapies and the capturing of the raw learning data arising in multiple modalities.
2) *Data pre-processing and annotation:* This refers to extracting useful data from the raw records, producing data traces in the required modality, performing data fusion by combining the traces, and adding the learning labels to the fused data to form labelled samples.
3) *Data processing, model building and evaluation:* This consists of standard ML procedures, including any necessary resampling, model building, training, testing, and performance evaluation.

The subsequent subsections will be more elaborate on each of the above stages and the involved procedures.

A. MULTIMODAL LEARNING DATA COLLECTION

We used the IoT-based system presented in Section III-C to carry out multimodal learning data collection at this stage.

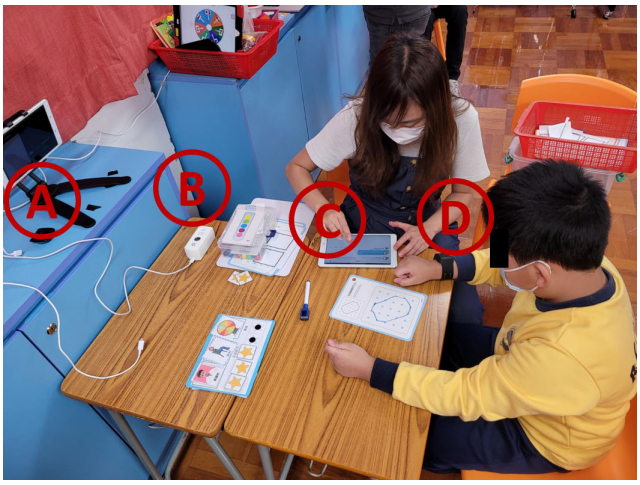

FIGURE 5. The setting of an ABA session in our study. (A) A screen display shows various real-time data related to the ABA session recommended by the 3I-Learning System. (B) An IoT sensors box contains environmental sensors. (C) An assessment interface records the therapist's task analysis results in real-time. (D) An Empatica E4 wristband collects the student's physiological and motion data.

A typical data collection scenario is shown in Fig. 5, in which a student and a therapist conduct ABA training using our system's perception layer equipment set. The multimodal learning data collected are described below.

1) EDUCATIONAL DATA

The educational data include demographic or contextual information such as the School (integer values representing each of the participating schools), the Student (distinct integer values representing the participants anonymously), and the Task domain (integer values representing the domain of the behavioural task performed). These data are encoded as three categorical variables: school, student, and task domain.

2) ENVIRONMENTAL DATA

The environmental data include the ambient carbon dioxide level (CO2), relative humidity (Humidity), ambient temperature (Temperature), and light intensity (Light) in lumens per square meter. These data were collected by a set of environmental sensors installed in an IoT sensor box (Fig. 2). All measurements were made at 1 Hz.

3) PHYSIOLOGICAL DATA

The physiological data include blood volume pulse (BVP) collected by a photoplethysmography sensor at 64 Hz, the inter-beat interval (IBI) time derived from BVP, galvanic skin response (GSR) collected by the electrodermal activity sensor at 4 Hz, and the participant's skin temperature (Skin Temperature) measured by the optical thermometer at 4 Hz. The collected data reflected the students' physiological condition during the ABA sessions in real time.

4) MOTION DATA

The motion data capture the participant's body movement in the left (+ve) and right (−ve), up (+ve) and down (−ve), and

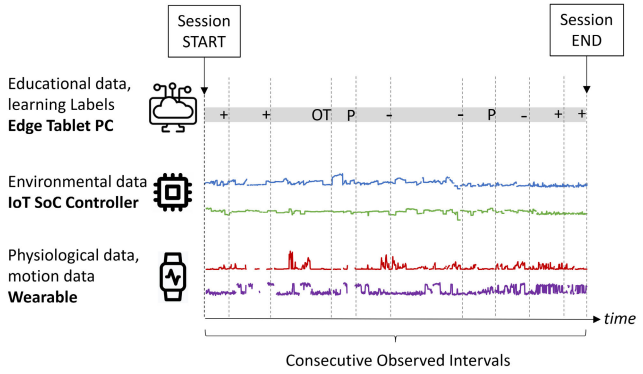

FIGURE 6. Data collection and features generation.

front (+ve) and rear (−ve) directions simultaneously. They were called Acceleration X, Y, and Z, respectively. These values are detected by the MEMS-type 3-axis accelerometer on the BLE wearable and are recorded from the participant's wristband-wearing hand at 32Hz.

5) PERFORMANCE DATA

Following the criteria of ABA assessment, each behaviour response observed during the therapy sessions was assessed as plus (''+''), minus (''−''), prompt (''P''), or off task (''OT''). The therapists input these assessment markers based on their subjective judgment. The inter-rater reliability alpha of this study is 0.96.

B. DATA PRE-PROCESSING AND ANNOTATION

At this stage, the raw data are pre-processed, fused, and annotated for further analysis. We also perform statistical analyses to determine if further data pre-processing procedures, such as data standardisation, are necessary.

1) FEATURE ENGINEERING

The environmental data (collected by the SoC controller), physiological and motion data (collected by the wristband), and performance data (collected by the edge tablet PC) collected were stored in three separate JSON files, with timestamps being appended to every datum. The raw sensor data streams are treated as time series signals. As illustrated in Fig. 6, the session averages of each of the data streams are generated as features.

2) LEARNING LABELS GENERATION AND ANNOTATION

We use the ABA therapy **correct rate of response (CR%)** as the student's performance indicator. The CR% is defined as the number of correct responses (#''+'') divided by the total number of response opportunities within an observed interval. Since every response opportunity was given an assessment marker, therefore we have the following:

$$\text{CR\%} = \left(\frac{\#\text{``+''}}{\#\text{``+''} + \#\text{``-''} + \#\text{``P''} + \#\text{``OT''}} \right) \times 100\%. \quad (1)$$

We compute our outcome variable according to the mastery criteria in the behaviour analysis literature [35]. The outcome is a binary variable that indicates whether a behavioural skill mastered in the therapy session can be maintained later. It is jointly determined by the therapy session CR% and the corresponding maintenance probe session CR%, where:

1) *Behavior Change* = 1 if therapy session CR% $\geq$ 90 and probes session CR% $\geq$ 90; and
2) *Behavior Change* = 0 otherwise.

The values of *Behavior Change* are used as the learning labels. In the end, the environmental, physiological, and motion features are fused with the educational data and learning labels to become annotated multimodal samples.

3) STATISTICAL ANALYSES

We use RStudio (with R v.4.2.2) and IBM SPSS v.27 to perform statistical analyses. First, we undergo outliers identification, removal, and missing value replacement. We then obtain the descriptive statistics (mean, standard deviation, skewness, and Kurtosis value) to gain a preliminary understanding of our data distribution. We also explore the statistical relationship between different features by correlation analysis and examine whether specific handling of statistical issues such as multicollinearity is needed.

We run binomial logistic regression analyses to learn the predictive relations of various features and the outcome. We also compare models to investigate whether the IoT sensor data can improve the predictive performance of a model containing only educational data. Since our features span a wide range of values (e.g., mean equals 770.60 for CO2 and 0.59 for IBI), we perform feature standardisation to improve the ML algorithms' potential performance. For a variable x, we obtain its standardised Z-score x' by the formula where

$$x' = \frac{x - \text{mean}(x)}{\text{standard deviation}(x)}. \quad (2)$$

C. MODELING BUILDING AND EVALUATION

Given the annotated samples resulting from the data pre-processing stage, we carry out standard ML procedures, such as class balancing, training, cross-validation, and testing, to produce our predictive model.

1) DATA PROCESSING PIPELINE

The data pipeline of our ML procedures is presented in Fig. 7. Firstly, we divide our samples (N = 1,130) into training and test sets in an 80% to 20% ratio. The test samples (n = 226) are held out and used exclusively for the testing phase. Various resampling methods are then applied to the training set (n = 904). Validation sets have been randomly extracted from the resampled training set to assess the training model's convergence. Lastly, the held-out training samples are used to evaluate the optimised model. We evaluate all trained models with metrics, including accuracy, precision, recall, and F-1 scores. The most optimum predictive model for the data is selected.

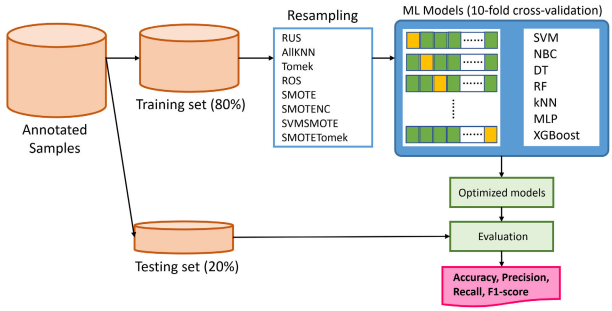

FIGURE 7. Data pipeline of the current study.

TABLE 2. Resampling methods and algorithms used in the current study.

Resampling Method	Algorithms
Downsampling	AllKNN RandomUnderSampler TomekLinks
Upsampling	RandomOverSampler SMOTE, SMOTE-NC SVMSMOTE
Hybrid	SMOTETomek

TABLE 3. Descriptive statistics of the feature variables (N = 1,130).

Feature (Unit / Label)	Mean	SD	Skewness†	Kurtosis‡
CO_2 Level (PPM)	770.60	196.42	0.52	-0.24
Humidity (RH)	53.37	9.56	-0.43	-0.30
Temperature (°C)	21.46	1.40	0.21	1.64
Light intensity (Lux)	332.26	129.15	-0.03	-0.40
BVP (n.a.)	-0.01	4.44	-2.85	139.42
IBI (s)	0.59	0.06	1.39	6.90
GSR (μS)	1.72	2.73	3.09	11.91
Skin temperature (°C)	33.13	2.23	-1.56	4.15
Acceleration X (ms^{-2})	-17.45	35.84	0.87	-0.80
Acceleration Y (ms^{-2})	-1.28	17.69	0.19	2.92
Acceleration Z (ms^{-2})	15.78	15.46	-0.62	1.39

† Standard error = 0.07.
‡ Standard error = 0.15.

2) RESAMPLING AND CLASS BALANCING

Uneven class balance is a frequent problem in real-world ML practice. Prior to our ML modelling process, we examine our statistical analysis result to detect any uneven class balance within our dataset. We then apply any necessary data augmentation techniques to enhance the class balance of our training data. We use APIs from the Python `imbalanced-learn` toolbox to perform data resampling.

The imbalanced classes problem exists in our annotated samples, where the size of negative and positive samples are $n_0 = 951$ and $n_1 = 179$, respectively. Therefore, we apply those standard resampling methods and algorithms listed in Table 2 to augment our training dataset.

3) ML MODELS BUILDING AND EVALUATION

We employed a range of well-established classifiers to construct the predictive models. Specifically, we used the k-nearest neighbours (kNN), decision tree (DT), random forest (RF), Naive Bayes classifier (NBC), multi-layer perceptrons (MLP), support vector machine (SVM), and XGBoost algorithms to build these classifiers. In addition, we also utilised a deep neural network (DNN) as a more advanced ML technique for classification.

To ensure that our models were reliable and accurate, we followed rigorous training, validation, and testing procedures in standard ML practice. We used the data pipeline described in Fig. 7 to split the data into training and testing sets. The training set was used to train the classifier models, while the validation set was used to tune their hyperparameters and prevent overfitting. Finally, the testing set was used to evaluate the performance of the models on unseen data. This approach allowed us to identify the most suitable ML algorithm for our specific problem and to optimise its performance through careful hyperparameter tuning.

TABLE 4. Binary Logistic Regression Results (N = 1,130).

	Model 1	Model 2	Model 3
Continuous Variables	*B (SE)*	*B (SE)*	*B (SE)*
CO2	-0.08 (0.10)	–	0.13 (0.15)
Humidity	0.18 (0.11)	–	-0.21 (0.15)
Temperature	-0.18 (0.12)	–	-0.20 (0.20)
Light	0.15 (0.10)	–	-0.40 (0.20)*
BVP	-0.01 (0.09)	–	0.00 (0.13)
IBI	-0.17 (0.10)	–	-0.00 (0.15)
GSR	0.13 (0.08)	–	0.28 (0.15)*
Skin Temperature	0.25 (0.11)*	–	0.58 (0.22)**
Acceleration X	0.28 (0.08)**	–	0.26 (0.14)*
Acceleration Y	0.11 (0.09)	–	0.09 (0.13)
Acceleration Z	-0.31 (0.09)**	–	-0.13 (0.12)
Categorical Variables	–	*Wald*	*Wald*
School	–	0.10	0.41
Student	–	95.03***	91.77***
Task Domain	–	35.48***	26.88***
Model Summary			
-2 Log Likelihood Ratio	868.40	647.54	618.48
Cox & Snell's R^2	0.04	0.21	0.23
Nagelkerke's R^2	0.07	0.38	0.42

*p< .05, **p< .01, ***p< .001
Dependent variable: Behaviour Change

V. RESULTS

A. DESCRIPTIVE STATISTICS

Descriptive statistics of each of the features are listed in Table 3. Both skewness and Kurtosis values of our variables indicate that the data in most of our variables were not normally distributed. Therefore, non-parametric statistical methods that do not assume data normality will be used in our subsequent analyses.

B. BINARY LOGISTIC REGRESSION ANALYSES

We performed a series of binary logistic regression analyses to study the predictive effects of the measured variables on participants' behaviour change. We compared the predictive ability among statistical models with (1) IoT sensor data

TABLE 5. Spearman's Rho Partial Correlation Matrix of the Study Variables Controlled by Student, School, and Task Domain (N = 1,130).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. CO2	–									
2. Humidity	0.11***	–								
3. Temperature	-0.11***	-0.31***	–							
4. Light	0.16***	-0.29***	0.16***	–						
5. BVP	0.00	-0.02	0.02	0.05	–					
6. IBI	-0.01	-0.01	-0.13***	0.04	-0.01	–				
7. GSR	0.11***	0.01	0.20***	0.17***	-0.05	-0.09**	–			
8. Skin Temperature	0.08**	-0.07*	0.33***	0.11***	-0.00	-0.22***	0.18***	–		
9. Acceleration (X Direction)	-0.04	0.00	0.14	-0.02	-0.05	0.12***	0.10**	0.06*	–	
10. Acceleration (Y Direction)	-0.10**	0.05	-0.01	-0.17***	0.04	-0.08*	-0.06	0.03	-0.03	–
11. Acceleration (Z Direction)	-0.01	-0.12***	0.07*	0.16***	0.03	-0.06*	0.11***	0.24***	0.05	-0.02

*p< .05, **p< .01, ***p< .001

(model 1), traditional education data (model 2), and a combination of both IoT sensor data and traditional education data (model 3). The resulting models are specified in Table 4.

Omnibus tests show that all of the models are statistically significant, where $\chi^2(11) = 43.59$, $p < .001$ (model 1), $\chi^2(29) = 264.45$, $p < .001$ (model 2), and $\chi^2(41) = 293.51$, $p < .001$ (model 3). Model 3 significantly improves over model 2 ($\Delta\chi^2 = 29.06$, $\Delta df = 11, p < .01$), and explains most of the variation of the outcome (Nagelkerke's $R^2 = 41.5\%$). Hosmer-Lemeshow tests indicate that the data fit both model 1 (χ^2=10.92, df = 8, p = 0.21) and model 3 (χ^2=12.57, df = 8, p = 0.13) but not model 2 (χ^2=16.88, df = 8, $p = 0.03$).

C. CORRELATION ANALYSES

Zero-order correlation matrix amongst the non-categorical variables is provided in Table 5. Partial correlation with the School, Student, and Task domain variables controlled is performed. Correlations between the predictors are found. For example, the CO2 level is significantly and positively correlated to GSR ($\rho = 0.14$, $p < 0.001$) and negatively correlated to Acceleration Y ($\rho = -0.11$, $p < 0.001$). Humidity negatively correlates to Acceleration Z ($\rho = -0.13$, $p < 0.001$). Temperature is significantly correlated to most of the physiological (IBI ρ = -0.13, $p < 0.001$; GSR $\rho = 0.21$, $p < 0.001$; Skin temperature $\rho = 0.33$, $p < 0.001$) and motion predictors (Acceleration X $\rho = 0.14$, $p < 0.001$ and Acceleration Y $\rho = 0.08$, $p < 0.01$); except the BVP and Acceleration in the Y direction. While Light is significantly correlated to GSR ($\rho = 0.18$, $p < 0.001$), Skin temperature ($\rho = 0.11$, $p < 0.001$), Acceleration X ($\rho = -0.17$, $p < 0.001$) and Acceleration Z ($\rho = 0.17$, $p < 0.001$). Nevertheless, no strong correlations ($\rho \geq 0.40$) that might indicate multicollinearity is found among the predictors.

D. ML MODELS EVALUATION

1) MODEL OPTIMIZATION AND CROSS-VALIDATION

We used `GridSearchCV` of the `scikit-learn 1.2.2` Python open-source library to tune the hyper-parameters of our classifiers. The grid search method exhaustively generates candidate values specified in a custom-defined hyperparameter space and returns the values for the best cross-validation score. We used 10-fold cross-validation to verify the generalisation ability of the resulting classifiers. We repeat the above process for each of the resampling

TABLE 6. Accuracy, precision, recall, and F1 score of the optimised SVM classifiers for each resampling algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
RUS	69.47%	30.77%	82.35%	44.80%
AllKNN	76.11%	32.76%	55.88%	41.30%
Tomek	82.74%	42.42%	41.18%	41.79%
None	82.74%	42.42%	41.18%	41.79%
ROS	85.84%	53.12%	50.00%	51.52%
SMOTE	77.43%	38.67%	85.29%	53.21%
SMOTENC	68.58%	31.68%	94.12%	47.41%
SVMSMOTE	90.71%	63.83%	88.24%	74.07%
SMOTETomek	**92.04%**	**68.18%**	**88.24%**	**76.92%**

TABLE 7. Accuracy, precision, recall, and F1 score of the optimised Naive Bayes classifier for each resampling algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
RUS	65.93%	23.46%	55.88%	33.04%
AllKNN	68.58%	24.66%	52.94%	33.64%
Tomek	72.12%	28.99%	58.82%	38.83%
Original	74.78%	33.33%	67.65%	44.66%
ROS	57.96%	21.50%	67.65%	32.62%
SMOTE	63.27%	23.66%	64.71%	34.65%
SMOTENC	71.24%	20.75%	32.35%	25.29%
SVMSMOTE	71.24%	28.17%	58.82%	38.10%
SMOTETomek	**75.66%**	**34.33%**	**67.65%**	**45.54%**

TABLE 8. Accuracy, precision, recall, and F1 score of the optimised Decision Tree classifier for each resampling algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
RUS	56.64%	23.33%	82.35%	36.36%
AllKNN	77.88%	35.71%	58.82%	44.44%
Tomek	81.42%	34.62%	26.47%	30.00%
Original	77.43%	28.21%	32.35%	30.14%
ROS	80.53%	37.50%	44.12%	40.54%
SMOTE	**95.58%**	**80.00%**	**94.12%**	**86.49%**
SMOTENC	92.92%	73.68%	82.35%	77.78%
SVMSMOTE	94.69%	82.35%	82.35%	82.35%
SMOTETomek	90.27%	63.64%	82.35%	71.79%

TABLE 9. Accuracy, precision, recall, and F1 score of the optimised Random Forest classifier for each resampling algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
RUS	67.26%	26.19%	64.71%	37.29%
AllKNN	79.65%	37.50%	52.94%	43.90%
Tomek	86.73%	60.00%	35.29%	44.44%
Original	87.17%	66.67%	29.41%	40.82%
ROS	87.61%	66.67%	35.29%	46.15%
SMOTE	**97.79%**	**96.77%**	**88.24%**	**92.31%**
SMOTENC	64.60%	29.46%	97.06%	45.21%
SVMSMOTE	96.46%	93.33%	82.35%	87.50%
SMOTETomek	93.81%	79.41%	79.41%	79.41%

TABLE 10. Accuracy, precision, recall, and F1 score of the optimised k-Nearest Neigbhour classifier for each resampling algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
RUS	96.02%	83.78%	91.18%	87.32%
AllKNN	**96.02%**	**82.05%**	**94.12%**	**87.67%**
Tomek	76.11%	27.27%	35.29%	30.77%
Original	94.25%	75.61%	91.18%	82.67%
ROS	83.63%	45.95%	50.00%	47.89%
SMOTE	96.02%	87.88%	85.29%	86.57%
SMOTENC	65.49%	30.00%	97.06%	45.83%
SVMSMOTE	95.13%	79.49%	91.18%	84.93%
SMOTETomek	92.48%	68.89%	91.18%	78.48%

TABLE 11. Accuracy, precision, recall, and F1 score of the optimised Multi-layer Perceptron classifier for each resampling algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
RUS	66.22%	24.71%	65.62%	35.90%
AllKNN	75.23%	26.53%	40.62%	32.10%
Tomek	84.23%	38.46%	15.62%	22.22%
None	86.04%	54.55%	18.75%	27.91%
ROS	81.08%	40.00%	62.50%	48.78%
SMOTE	88.74%	60.61%	62.50%	61.54%
SMOTENC	74.77%	35.00%	87.85%	50.00%
SVMSmote	89.19%	62.50%	62.50%	62.50%
SMOTETomek	**88.29%**	**56.82%**	**78.12%**	**65.79%**

methods. In order to emphasise the successful classification of minority cases, classifiers with the highest F1 scores (calculated from precision and recall) are selected.

2) CLASSIFIERS MODEL PERFORMANCE AND EVALUATION

The details of the performance of our classifiers are given in Tables 6 to 11. Accuracy, precision, recall, and F1 scores for models using each classification method and resampling algorithm introduced above are listed.

Our results show that the classifiers can predict the achievement of behaviour change with high accuracy in general. In particular, the RF classifier with SMOTE upsampling achieves the highest accuracy of 97.79%. The same RF classifier also achieves the highest precision (96.77%) and F1 score (92.31%). At the same time, the XGBoost classifier (with SMOTETomek hybrid resampling) gives the highest recall (97.06%). In general, the kNN classifier has the best performance in terms of the averages of all matrices.

TABLE 12. Accuracy, precision, recall, and F1 score of the optimised XGBoost classifier for each resampling algorithm.

Algorithm	Accuracy	Precision	Recall	F1 Score
RUS	62.39%	23.16%	64.71%	34.11%
AllKNN	74.34%	30.65%	55.88%	39.58%
Tomek	86.73%	57.14%	47.06%	51.61%
None	87.61%	62.50%	44.12%	51.72%
ROS	83.19%	44.44%	47.06%	45.71%
SMOTE	92.92%	75.00%	79.41%	77.14%
SMOTENC	71.24%	33.33%	91.18%	48.82%
SVMSMOTE	96.90%	90.91%	88.24%	89.55%
SMOTETomek	**97.35%**	**86.84%**	**97.06%**	**91.67%**

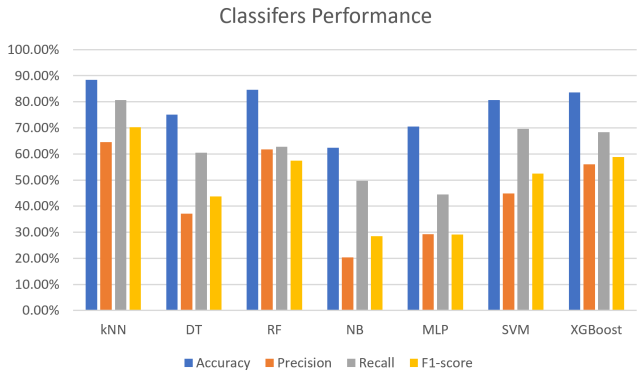

FIGURE 8. Mean accuracy, precision, recall, and F1 score by classifiers.

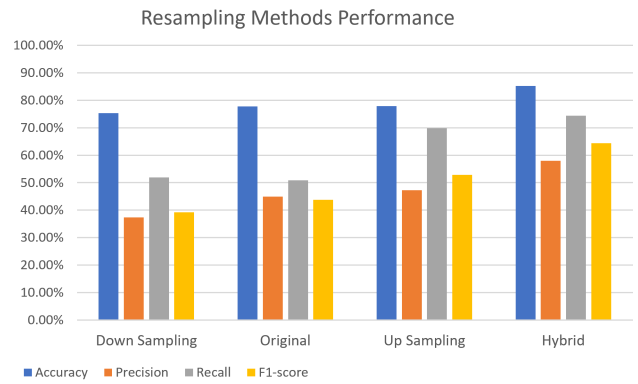

FIGURE 9. Mean accuracy, precision, recall, and F1 score by resampling methods.

Its average accuracy, precision, recall, and F1 score are 88.35%, 64.55%, 80.72%, and 70.24%, respectively. However, the NB classifier has the poorest performance with average accuracy, precision, and F1 score equal to 62.44%, 20.39%, and 28.45%, respectively. In comparison, the MLP classifier has the lowest recall (44.44%).

The mean performance scores by classifiers and by resampling methods are shown in Fig. 8 and 9, respectively. It is shown that, on average, kNN classifiers have the best performance. NB classifiers have the lowest accuracy, precision, and F1 scores, while MLP has the lowest recall and F1 scores. For resampling methods, the hybrid method produces the best performance. Downsampling gives the worst performance in

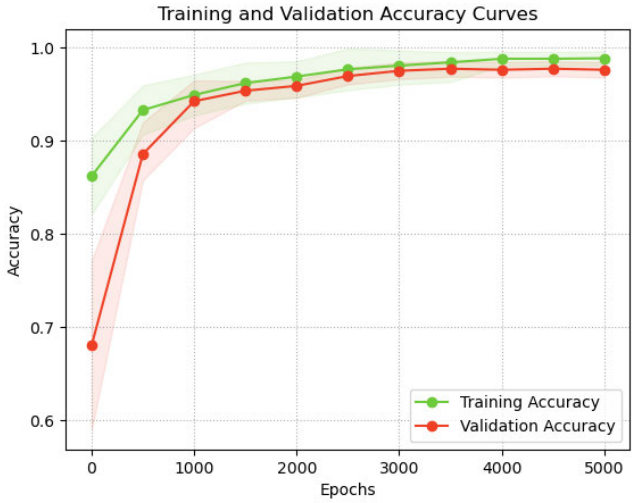

FIGURE 10. Training and validation accuracy by epochs.

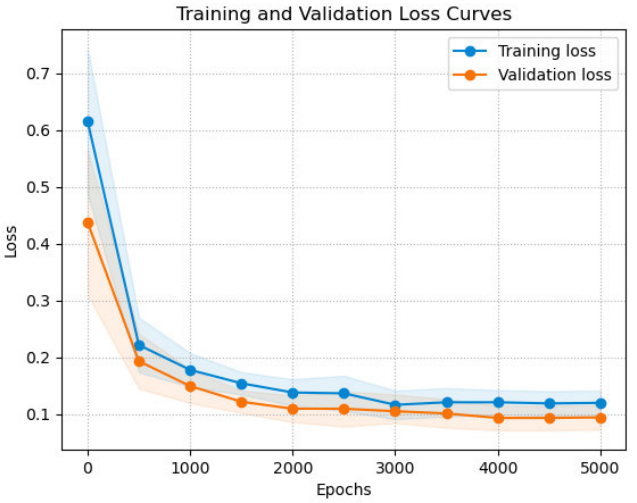

FIGURE 11. Training and validation loss by epochs.

TABLE 13. Optimum Performance for DNN.

Resampling Method	Accuracy	Precision	Recall	F1 Score
None	97.75%	93.55%	90.62%	92.06%

terms of accuracy, precision, and F1 score, while the performance in terms of recall is the poorest when no resampling method is applied (i.e., original).

3) DEEP NEURAL NETWORK BUILDING, TRAINING, AND EVALUATION

We used `TensorFlow 2.11.0` to build our deep neural network (DNN) and ran our program on a GPU (NVIDIA RTX A2000 12GB). We established a DNN with 46 input nodes (for the predictive variables) and two output nodes (for the two classes). Our best-performed DNN model did not use any resampling, and the resulting DNN has four hidden layers, each having 32 nodes, respectively. We use hyperbolic tangent (tanh) as the activation function. We have run 5 000 epochs with a batch size of 118, a dropout rate of 0.1, a momentum of 0.92, and an initial learning rate of 0.002.

We present the performance of our DNN using two sets of learning curves. The training and validation accuracy curve (Fig. 10) shows an increasing trend in our model's training accuracy. An increasing trend can also be observed in the validation accuracy curve, but there is no further improvement in the validation accuracy after 3 000 epochs. The training and validation loss curves (Fig. 11) show that both the training and validation loss of our model decreased over epochs. The evaluation metrics are given in Table 13.

4) OVERALL PERFORMANCE

The Precision vs Recall scattered plot of all classifiers and the DNN in the current study is provided in Fig. 12. The best-performed classifier-resampling method combinations

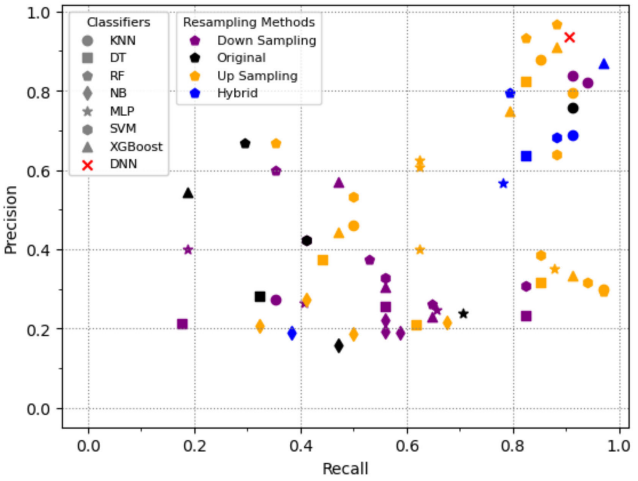

FIGURE 12. Precision vs Recall scattered plot of all classifier-resampling method combinations.

are located in the upper left-hand corner of the plot. The worse ones are located in the lower left-hand corner.

VI. DISCUSSIONS

A. STATISTICAL CHARACTERISTICS AND RELATIONS BETWEEN PREDICTORS

Most of our predictor variables (Table 4) are not normally distributed. Therefore, we used non-parametric statistical methods in our correlation and regression analyses. We also standardised our data before feeding them to the ML algorithms and DNN. These precautions can enhance the generalisation and performance of our prediction models.

Spearman's rho partial correlation analyses show statistically significant correlations between the environmental and physiological predictors (Table 3). For example, the CO2 level of the classroom is significantly and positively correlated to the participants' GSR ($\rho = 0.11, p < 0.001$). The classroom temperature is significantly correlated to

participants' skin temperature ($\rho = 0.33, p < 0.001$), GSR ($\rho = 0.20, p < 0.001$), and IBI ($\rho = -0.13, p < 0.001$). Besides, classroom light intensity is significantly correlated to participants' GSR ($\rho = 0.17, p < 0.001$) and skin temperature ($\rho = 0.11, p < 0.001$).

Our results show that SEN students' motions significantly correlate to the ambient learning environment. As reflected from Acceleration Y (upward vs downward motions), the classroom CO2 level is negatively correlated to participants' upward position ($\rho = -0.10, p < 0.01$). Acceleration Z (forward vs backward motions) is negatively correlated to humidity ($\rho = -0.12, p < 0.001$). Besides, light intensity is positively correlated to downward ($\rho = -0.17, p < 0.001$) and forward motions ($\rho = 0.16, p < 0.001$).

No strong correlations ($\rho \geq 0.4$) were found among our predictors. Besides, the Variance Inflation Factor (VIF) values of our predictors range from 1.00 (BVP) to 1.49 (temperature). Therefore, we confirm that our predictors are only moderately correlated, and our regression analyses are not likely to be subject to multicollinearity.

B. PREDICTION OF BEHAVIOR CHANGE IN SEN STUDENTS USING SENSORS AND WEARABLE DATA

Our binary logistic regression analysis results (Table 5) show that traditional educational data (model 2) can explain 38% (Nagelkerke's R^2) of the variance in participants' state of behaviour change. The percentage is further improved to 42% when sensors and wearable data are included (model 3). This result suggests that the inclusion of sensors and wearable data can better predict behaviour change than using traditional educational data alone.

Our full model shows statistically significant predictive relations of light intensity ($B = -0.40, SE = 0.20, p < 0.05$) and GSR ($B = 0.28, SE = 0.15, p < 0.05$) on behaviour change. These results align with existing literature in special education. The negative relation between excessive imminence from classroom lighting and discomfort in SEN students has been reported in the literature (e.g., [36]). Indeed, strong lighting should be avoided in classroom settings for children with ASD, including those with Asperger syndrome [37], because their neural system is often oversensitive to light sources. The induced pattern glare might affect their learning performance [38]. The identified positive predictive relation of GSR also aligns with findings from special education research [17], where SEN students who did not actively engage in the task were found to have low skin conductance levels. In particular, SEN students who could not maintain a constant attention level had significantly lower skin conductance levels (p.45). The relationship between GSR and social skills was further supported by a study on students with intellectual disabilities [39], in which a higher sympathetic activation was related to more skin conductance responses. Since skin conductance is directly involved in human emotional and behavioural regulation [40], our results suggest that GSR is a plausible feature for predicting SEN students' behavioural learning.

Besides, we find significant and positive predictive relations of skin temperature ($B = 0.58, SE = 0.22, p < 0.01$) on behaviour change. Our result aligns with a recent study [41], where typically developed students' skin temperature during learning increases significantly in high engagement over low engagement (p.279). However, developmental disabilities such as ASD may be associated with atypical autonomic nervous responses, making SEN students' skin temperature change response to anxiety less salient than the typically developed student groups [42]. Lastly, both Student and Task domain categorical variables significantly predict behaviour change ($p < 0.001$). This result suggests that student-oriented personalisation and learning-task customisation of MMLA models are necessary for accurate prediction.

C. MMLA AND PREDICTIVE MODELING IN BEHAVIOR CHANGE FOR SEN STUDENTS

We have developed eight predictive models (seven ML-based classifiers and a DNN) and evaluated their performance. The overall performance in terms of specificity (precision) and sensitivity (recall) is summarised in Fig. 11. Up-sampling methods generally outperform the others in precision, while hybrid and down-sampling methods give the best recall. In particular, the XGBoost-Hybrid and RF-Up Sampling combinations give the best performance in terms of recall and precision, respectively. Meanwhile, our optimised DNN falls onto the good performance cluster (the red cross). Classifiers with MLP-Down Sampling combination and all NB classifiers generally perform worse than the rest of the combinations. Table 14 lists our optimised classifiers by their accuracy scores and provides their pros and cons concerning the context of MMLA for SEN.

We compare the prediction performance of our models with other existing MMLA models. Table 15 lists the accuracy, precision, recall, and F1 score presented in each work. The performance of our models is comparable to those in the previous works. Our RF classifier outperforms the existing works in terms of all metrics. While our RF, XGBoost, and DNN are equally well performed in terms of F1 scores. However, it is also noted that there are diverse prediction targets and data modalities among these MMLA models. Therefore, Table 15 could only be used as a reference. Lastly, by considering both performance and the pros and cons among our eight models, we select the XGBoost classifier for behaviour change prediction.

VII. LIMITATIONS AND FUTURE WORK

We are aware of several limitations of our current work:

- Our prediction target is a binary output, which limits the available information regarding students' ABA learning for the teachers and therapists.
- The current data collection system works in a one-to-one therapist-to-student setting. While in the daily special education context, classroom teaching

TABLE 14. Ranking of optimised classifiers in the current study by accuracy scores; and their pros and cons with MMLA for SEN considerations.

Rank (Accuracy)	Classifier	Pros	Cons
1. (97.79%)	RF	High performance. Capable of handling non-linear multimodal educational data.	Higher training time when compared to DT. It may not be suitable for edge devices in SEN classrooms.
2. (97.75%)	DNN	High performance. Suitable for prediction tasks in SEN that require high accuracy.	Lack of explainability and interpretability to inform decision-making. Computationally expensive.
3. (97.35%)	XGBoost	Fast and high performance. Can handle missing values.	Sensitive to outliers, whereas outliers often appear in MMLA for SEN.
4. (96.02%)	kNN	Simple to implement. Work well with normalised and standardised features.	It becomes significantly slower when the number of samples increases.
5. (95.58%)	DT	Suitable for small features size. Fast to construct. Explainable and therefore suitable for MMLA.	Easily overfit, large trees are difficult to interpret and adapt to new rules.
6. (92.04%)	SVM	Memory-efficient and suitable for edge ML. Suitable for special schools with limited resources.	It does not perform well with noisy data, which is often the case for physiological data.
7. (88.29%)	MLP	Can model complex non-linear input-output relationships. Suitable for physiological data.	Performance is highly dependent on the data quality. Not suitable for edge devices in SEN classrooms.
8. (75.66%)	NB	Fast and does not require much training data. Suitable for SEN, where the sample size is often small.	Assumes independence among features which is difficult to achieve in MMLA for SEN.

TABLE 15. Comparison of the predictive performance of MMLA models.

Authors	Model	Accuracy	Precision	Recall	F1 score	Prediction Target	Data Modality
Ma *et al.* [26]	SVM	0.81	0.55	0.77	0.63	Impasses	linguistic, audio, video
Ma *et al.* [26]	MLP	0.84	0.56	0.79	0.65	Impasses	linguistic, audio, video
Kawamura *et al.* [27]	SVM	N/A	N/A	N/A	0.75	Wakefulness	physiological, motion, face
Kawamura *et al.* [27]	RF	N/A	N/A	N/A	0.71	Wakefulness	physiological, seat pressure, face
Azcona *et al.* [28]	kNN	0.77	0.68	0.88	0.77	Course performance	demographics, education, motion
Azcona *et al.* [28]	SVM	0.68	0.86	0.64	0.73	Course performance	demographics, education, motion
Azcona *et al.* [28]	DT	0.48	**0.94**	0.55	0.69	Course performance	demographics, education, motion
Sharma *et al.* [29]	SVM	N/A	0.81	0.78	0.80	Learning effort	physiological, logs, survey
Sharma *et al.* [29]	RF	N/A	0.79	0.73	0.76	Learning effort	physiological, logs, survey
Emerson *et al.* [30]	LR	0.61	N/A	N/A	N/A	Game-based learning	logs, face, gaze, survey
Mangaroska *et al.* [31]	RF	N/A	0.86	0.84	0.85	Debugging expertise	logs, face, gaze, physiological
Spikol *et al.* [32]	NB	0.80	N/A	N/A	N/A	Artefact quality	education, audio, face, motion
Spikol *et al.* [32]	SVM	0.75	N/A	N/A	N/A	Artefact quality	education, audio, face, motion
Current Work	kNN	0.96	0.82	**0.94**	0.88	Behaviour change	education, physiological, motion, environmental
Current Work	DT	0.96	0.82	0.82	0.82	Behaviour change	education, physiological, motion, environmental
Current Work	RF	**0.98**	**0.97**	0.88	**0.92**	Behaviour change	education, physiological, motion, environmental
Current Work	NB	0.76	0.27	0.41	0.33	Behaviour change	education, physiological, motion, environmental
Current Work	MLP	0.88	0.57	0.78	0.66	Behaviour change	education, physiological, motion, environmental
Current Work	SVM	0.92	0.68	0.88	0.77	Behaviour change	education, physiological, motion, environmental
Current Work	XGBoost	**0.97**	0.87	**0.97**	**0.92**	Behaviour change	education, physiological, motion, environmental
Current Work	DNN	**0.98**	**0.93**	**0.91**	**0.92**	Behaviour change	education, physiological, motion, environmental

Notes. SVM = Support vector machine, kNN = k-nearest neighbours, DT = Decision tree, RF = Random forest, LR = linear regression, NBC = Naive Bayesian classifier, MLP = Multi-layer Perceptrons. The best three results under each performance metric are **bolded**.

is often conducted in one-to-few or one-to-many manners.
- The measurement hardware in the current study is costly. For example, Empatica E4 wristbands were used, while an E4 wristband can cost more than a thousand US dollars.

The following future works are proposed based on the existing achievements of the current work:

- The current study affirms a statistically significant predictive relation between multimodal educational data and SEN students' behavioural change. As a future work, prediction models with multiple outcomes and at multiple levels can be developed. In this way, the system can provide more faceted information about SEN students' behavioural change to ABA practitioners.
- The current study demonstrates the feasibility of an MMLA predictive model for one-to-one ABA therapy. The work can be expanded to one-to-few or one-to-many settings.
- The current study has revealed a number of environmental and physiological variables that can predict behaviour change. In the future, alternative or new measurement devices at a lower cost can be explored or developed for SEN students.

VIII. CONCLUSION

In this paper, we applied MMLA to predict behaviour change in SEN students participating in ABA therapies. A novel MMLA approach for the prediction of SEN students' behaviour change achievement in ABA therapy is presented.

We introduced IoT sensors data, including ambient environmental measurements (namely CO2 level, humidity, light intensity, and temperature), physiological measurements (namely IBI, BVP, GSR, and skin temperature), and motion measurements (accelerometer values in X, Y, and Z directions) to develop statistical models for ABA therapy. We also apply ML and DNN techniques to predict SEN students' behaviour change.

We studied the statistical characteristics of the multimodal educational data and found that most of our data are not normally distributed. Significant correlations between the variables had been identified, but the problem of multicollinearity did not exist in our variables. We further showed that sensors and wearable data could significantly enhance the prediction of SEN students' behaviour change achievement. Various ML algorithms and a DNN were built, optimised, and evaluated. Our results demonstrated that ML (including deep learning) could be applied to MMLA for predicting SEN students' behaviour change. While the performance of our classifiers and DNN surpass most of the existing MMLA models. However, we also observed variations in the prediction targets among the compared models.

Promoting positive behaviours in SEN students is important for their personal and social development. At the same time, ABA therapy is an effective intervention approach that aims at behaviour change in this population group. The learning environment and the learner physiology conditions during ABA therapy sessions are essential for understanding behaviour skills acquisition and their effect on subsequent behaviour change. The current study has affirmed the predictive relations between the learning environment, learner physiology, and the learning outcome in ABA therapy. A number of limitations and necessary future works are also presented. Overall, our work echoes the growing demands in applying ML to the learning and education of those with brain and developmental disorders [43].

ACKNOWLEDGMENT

The authors would like to thank Christine Chan, Fiona Tsang, Albert Hui, Dennis Lee, Andrew Sze, Kangzhong Wang, Gary Lam, and Anthony Ng, and also would like to thank Bridge Academy, Caritas Resurrection School, and The Jockey Club Hong Chi School for their support. Rosanna Yuen-Yan Chan is a Principal Investigator of the Centre for Perceptual and Interactive Intelligence (CPII) under the InnoHK.

REFERENCES

[1] P. A. Alberto and A. C. Troutman, *Applied Behavior Analysis for Teachers*, 9th ed. Upper Saddle River, NJ, USA: Pearson, 2013.

[2] B. S. Abrahams and D. H. Geschwind, "Advances in autism genetics: On the threshold of a new neurobiology," *Nature Rev. Genet.*, vol. 9, no. 5, pp. 341–355, May 2008.

[3] L. Bassarath, "Conduct disorder: A biophysical review," *Can. J. Psychiatry*, vol. 46, no. 7, pp. 609–617, 2001.

[4] J. O. Cooper, T. E. Heron, and W. L. Heward, *Applied Behavior Analysis*, 3rd ed. Hoboken, NJ, USA: Pearson, 2020.

[5] R. Pennington, "Applied behavior analysis: A valuable partner in special education," *Teach. Except. Child.*, vol. 54, no. 4, pp. 315–317, 2022.

[6] F. J. Alves, E. A. De Carvalho, J. Aguilar, L. L. De Brito, and G. S. Bastos, "Applied behavior analysis for the treatment of autism: A systematic review of assistive technologies," *IEEE Access*, vol. 8, pp. 118664–118672, 2020.

[7] M. C. Buzzi, M. Buzzi, B. Rapisarda, C. Senette, and M. Tesconi, "Teaching low-functioning autistic children: ABCD SW," in *Proc. Eur. Conf. Technol. Enhanced Learn.* Berlin, Germany: Springer, 2013, pp. 43–56.

[8] V. Bartalesi, M. C. Buzzi, M. Buzzi, B. Leporini, and C. Senette, "An analytic tool for assessing learning in children with autism," in *Universal Access in Human-Computer Interaction, Universal Access to Information and Knowledge*, vol. 8514, C. Stephanidis and M. Antona, Eds. Cham, Switzerland: Springer, 2014.

[9] G. Presti, M. Scagnelli, M. Lombardo, M. Pozzi, and P. Moderato, "SMART SPACES: A backbone to manage ABA intervention in autism across settings and digital learning platforms," in *Proc. AIP Conf.*, vol. 2040, 2018, Art. no. 140002.

[10] G. Siemens and R. S. J. D. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," in *Proc. 2nd Int. Conf. Learn. Analytics Knowl.*, Apr. 2012, pp. 252–254.

[11] D. M. Baer, M. M. Wolf, and T. Risley, "Current dimensions of applied experimental analysis of behavior," *J. Appl. Behav. Anal.*, vol. 1, pp. 91–97, Jan. 1968.

[12] S. Vilcekova, L. Meciarova, E. K. Burdova, J. Katunska, D. Kosicanova, and S. Doroudiani, "Indoor environmental quality of classrooms and occupants' comfort in a special education school in Slovak Republic," *Building Environ.*, vol. 120, pp. 29–40, Aug. 2017.

[13] S. S. Ahmad, M. F. Shaari, R. Hashim, and S. Kariminia, "Conducive attributes of physical learning environment at preschool level for slow learners," *Proc.-Social Behav. Sci.*, vol. 201, pp. 110–120, Aug. 2015.

[14] M. M. Karima, "Evaluating impact of the lighting and acoustic environments on the learning development of children with cognitive disabilities in special education in UAE," M.S. thesis, Fac. Eng. IT, British Univ. Dubai, Dubai, UAE, 2017.

[15] A. Smith, *Accelerated Learning in the Classroom*. Stafford, U.K.: Network Continuum Education, 1996.

[16] A. Savan, "A study of the effect of background music on the behaviour and physiological responses of children with special educational needs," *Psychol. Educ. Rev.*, vol. 22, pp. 32–36, Feb. 1998.

[17] G. B. Werbach, "Skin conductance patterns among learning disabled students," Ph.D. dissertation, Dept. Special Educ., California State Univ., Fresno, CA, USA, 1979.

[18] P. Kosmas, A. Ioannou, and S. Retalis, "Moving bodies to moving minds: A study of the use of motion-based games in special education," *TechTrends*, vol. 62, no. 6, pp. 594–601, Nov. 2018.

[19] P. Blikstein and M. Worsley, "Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks," *J. Learn. Anal.*, vol. 3, no. 2, pp. 220–238, Sep. 2016.

[20] X. Ochoa and M. Worsley, "Augmenting learning analytics with multimodal sensory data," *J. Learn. Anal.*, vol. 3, no. 2, pp. 213–219, Sep. 2016.

[21] D. Di Mitri, J. Schneider, M. Specht, and H. Drachsler, "From signals to knowledge: A conceptual model for multimodal learning analytics," *J. Comput. Assist. Learn.*, vol. 34, no. 4, pp. 338–349, Aug. 2018.

[22] M. Kourakli, I. Altanis, S. Retalis, M. Boloudakis, D. Zbainos, and K. Antonopoulou, "Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using Kinect learning games," *Int. J. Child-Comput. Interact.*, vol. 11, pp. 28–39, Jan. 2017.

[23] P. Kosmas, A. Ioannou, and S. Retalis, "Using embodied learning technology to advance motor performance of children with special educational needs and motor impairments," in *Proc. Eur. Conf. Technol. Enhanced Learn.*, 2017, pp. 111–124.

[24] O. Rudovic, J. Lee, M. Dai, B. Schuller, and R. W. Picard, "Personalized machine learning for robot perception of affect and engagement in autism therapy," *Sci. Robot.*, vol. 3, no. 19, Jun. 2018, Art. no. eaao6760.

[25] M. S. Goodwin, C. A. Mazefsky, S. Ioannidis, D. Erdogmus, and M. Siegel, "Predicting aggression to others in youth with autism using a wearable biosensor," *Autism Res.*, vol. 12, no. 8, pp. 1286–1296, Aug. 2019.

[26] Y. Ma, M. Celepkolu, and K. E. Boyer, "Detecting impasse during collaborative problem solving with multimodal learning analytics," in *Proc. 12th Int. Learn. Analytics Knowl. Conf.*, Mar. 2022, pp. 45–55.

[27] R. Kawamura, S. Shirai, N. Takemura, M. Alizadeh, M. Cukurova, H. Takemura, and H. Nagahara, "Detecting drowsy learners at the wheel of e-Learning platforms with multimodal learning analytics," *IEEE Access*, vol. 9, pp. 115165–115174, 2021.
[28] D. Azcona, I. Hsiao, and A. F. Smeaton, "Personalizing computer science education by leveraging multimodal learning analytics," in *Proc. IEEE Frontiers Edu. Conf. (FIE)*, Oct. 2018, pp. 1–9.
[29] K. Sharma, Z. Papamitsiou, J. K. Olsen, and M. Giannakos, "Predicting learners' effortful behaviour in adaptive assessment using multimodal data," in *Proc. 10th Int. Conf. Learn. Analytics Knowl.*, Mar. 2020, pp. 480–489.
[30] A. Emerson, E. Cloude, R. Azevedo, and J. Lester, "Multimodal learning analytics for game-based learning," *Brit. J. Educ. Technol.*, vol. 51, no. 5, pp. 1505–1526, Jul. 2020.
[31] K. Mangaroska, K. Sharma, D. Gasevic, and M. Giannakos, "Multimodal learning analytics to inform learning design: Lessons learned from computing education," *J. Learn. Analytics*, vol. 7, no. 3, pp. 79–97, Dec. 2020.
[32] D. Spikol, E. Ruffaldi, G. Dabisias, and M. Cukurova, "Supervised machine learning in multimodal learning analytics for estimating success in project-based learning," *J. Comput. Assist. Learn.*, vol. 34, no. 4, pp. 366–377, Aug. 2018.
[33] G. Gripenberg, "Confidence intervals for partial rank correlations," *J. Amer. Stat. Assoc.*, vol. 87, no. 418, pp. 546–551, Jun. 1992.
[34] C. M. V. Wong, R. Y.-Y. Chan, Y. N. Yum, and K. Wang, "Internet of Things (IoT)-enhanced applied behavior analysis (ABA) for special education needs," *Sensors*, vol. 21, no. 19, p. 6693, Oct. 2021.
[35] C. B. McDougale, S. M. Richling, E. B. Longino, and S. A. O'Rourke, "Mastery criteria and maintenance: A descriptive analysis of applied research procedures," *Behav. Anal. Pract.*, vol. 13, no. 2, pp. 402–410, Jun. 2020.
[36] M. Winterbottom and A. Wilkins, "Lighting and discomfort in the classroom," *J. Environ. Psychol.*, vol. 29, no. 1, pp. 63–75, Mar. 2009.
[37] P. J. Katsioloudis and M. Jones, "Effects of light intensity on spatial visualization ability," *J. Technol. Stud.*, vol. 43, no. 1, pp. 2–13, Jan. 2017.
[38] B. Menzinger and R. Jackson, "The effect of light intensity and noise on the classroom behaviour of pupils with Asperger syndrome," *Support Learn.*, vol. 24, no. 4, pp. 170–175, Nov. 2009.
[39] P. Vos, P. De Cock, V. Munde, K. Petry, W. Van Den Noortgate, and B. Maes, "The tell-tale: What do heart rate; Skin temperature and skin conductance reveal about emotions of people with severe and profound intellectual disabilities?" *Res. Develop. Disabilities*, vol. 33, no. 4, pp. 1117–1127, Jul. 2012.
[40] G. I. Christopoulos, M. A. Uy, and W. J. Yap, "The body and the brain: Measuring skin conductance responses to understand the emotional experience," *Organizational Res. Methods*, vol. 22, no. 1, pp. 394–420, Jan. 2019.
[41] A. Ojha, H. Jebelli, and M. Sharifironizi, "Understanding students' engagement in learning emerging technologies of construction sector: Feasibility of wearable physiological sensing system-based monitoring," in *Proc. CSCE*, 2021, pp. 269–281.
[42] A. Kushki, E. Drumm, M. P. Mobarak, N. Tanel, A. Dupuis, T. Chau, and E. Anagnostou, "Investigating the autonomic nervous system response to anxiety in children with autism spectrum disorders," *PLoS ONE*, vol. 8, no. 4, Apr. 2013, Art. no. e59730.
[43] X. Chen, G. Cheng, F. L. Wang, X. Tao, H. Xie, and L. Xu, "Machine and cognitive intelligence for human health: Systematic review," *Brain Informat.*, vol. 9, no. 1, pp. 1–20, Dec. 2022.

ROSANNA YUEN-YAN CHAN (Senior Member, IEEE) received the B.Eng., M.Phil., and Ph.D. degrees in information engineering and the M.Ed. degree in educational psychology from The Chinese University of Hong Kong, Shatin, Hong Kong, in 1998, 2000, 2006, and 2009, respectively. She is currently a Senior Research Scientist with the Centre of Perceptual and Interaction Intelligence and an Adjunct Assistant Professor with the Department of Information Engineering, The Chinese University of Hong Kong. She is also the Founding Chair of the IEEE Education Society (EdSoc) Technical Committee on Learning Sciences and EdSoc Hong Kong Chapter. She also served as a Member-at-Large on EdSoc Board of Governors, from 2015 to 2021. She was a recipient of a number of IEEE awards in education, including the IEEE William E. Sayle Award for Achievement in Education and the IEEE EAB (Region 10) Major Educational Innovation Award.

CHUN MAN VICTOR WONG is currently pursuing the Ed.D. degree in special education with the Department of Special Education and Counseling, The Education University of Hong Kong. He is also an Industry Practitioner with the Department of Special Education and Counseling, The Education University of Hong Kong. He founded Bridge Academy, in 2014. He has also founded Bridge AI, a company that innovates in AI and machine learning for the e-learning of SEN students. He has received several funds from the Hong Kong Innovative and Technology Bureau (ITB) to develop educational systems that tailor to the individual needs of the SEN students.

YEN NA YUM received the Ph.D. degree in cognitive psychology from Tufts University. She is currently an Associate Professor and the Associate Head of the Department of Special Education and Counseling, The Education University of Hong Kong (EdUHK). Before joining EdUHK, she was a Postdoctoral Research Fellow with the Division of Speech and Hearing Sciences, The University of Hong Kong. Her research interests include bilingualism and second language learning, focusing on reading and writing processes across languages or writing systems. Her recent work has used the ERP method to examine the neural correlates of visual word processing and language learning among bilingual and multilingual populations.