IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 28 May 2024, accepted 18 July 2024, date of publication 29 July 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3434608

RESEARCH ARTICLE

Advancing Online Assessment Integrity: Integrated Misconduct Detection via Internet Protocol Analysis and Behavioral Classification

LESLIE CHING OW TIONG[1], YUNLI LEE[2], (Senior Member, IEEE), KAI LI LIM[3], AND HEEJEONG JASMINE LEE[4,5]

[1]Samsung Electronics Co., Ltd., Hwaseong-si, Gyeonggi-do 18448, Republic of Korea
[2]Research Centre For Human-Machine Collaboration (HUMAC), School of Engineering and Technology, Sunway University, Petaling Jaya, Selangor 46150, Malaysia
[3]Dow Centre for Sustainable Engineering Innovation, The University of Queensland, Brisbane, QLD 4072, Australia
[4]College of Information and Communication Engineering, Sungkyunkwan University, Jangan-gu, Suwon-si, Gyeonggi-do 16419, Republic of Korea
[5]SKAIChips Company Ltd., Gwonseon-gu, Suwon-si, Gyeonggi-do 16431, Republic of Korea

Corresponding author: HeeJeong Jasmine Lee (heejl@skku.edu)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by Korean Government through the Ministry of Science and ICT (MSIT) under Grant RS-2023-00250496, and in part by the National Research and Development Program through the National Research Foundation of Korea (NRF) funded by MSIT under Grant 2020M3H2A1076786.

ABSTRACT The rapid expansion of online learning has ushered in new educational opportunities, but concurrently introduced challenges in preserving academic integrity during assessments. This transition accentuates the need to address the heightened risks of e-cheating, encompassing various forms of academic dishonesty and potential misuse of advanced tools such as ChatGPT. To tackle these challenges, this paper proposes a pioneering multi-stage detection system that synergises geographical verification with behavioural analysis, offering a nuanced understanding of cheating behaviours in diverse online learning scenarios. The core innovation is an intelligent agent (multi-IA) incorporating an IP Detector for geographical verification and a deep learning-based Behavioural Monitor (known as DenseLSTM) for analysing response behaviours. This novel amalgamation enhances adaptability and accuracy in e-cheating detection, thereby fortifying the credibility and integrity of online assessments. Empirical analysis validates the multi-IA system's efficacy, demonstrating remarkable accuracy in classifying candidate behaviours and effectively discerning between normal and potentially fraudulent responses. These findings underscore the system's potential as an invaluable asset for educational institutions and educators in preserving the integrity of online assessments, and supporting the growth of online learning.

INDEX TERMS Online assessment integrity, e-cheating detection, intelligent agent in education, behavioral pattern analysis.

I. INTRODUCTION

As the digital transformation of education accelerates, online learning environments have become the new frontier in educational delivery. The Covid-19 pandemic has further accelerated this trend, as institutions decisively transitioned to remote education in response [1], [2], [3]. However, while online learning offers numerous benefits, it also presents novel challenges, particularly in maintaining academic integrity during synchronous online assessments. The remote and impersonal natures of online assessments, without the traditional oversight settings, raises concerns about increased risks of academic dishonesty, including cheating and collusion [4], [5], [6]. Our work specifically

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

focuses on addressing these challenges within the context of synchronous online assessments, where real-time evaluation is crucial for ensuring academic integrity.

In virtual learning spaces, maintaining honesty is particularly challenging. The impersonal setting of online assessments amplifies the risk of dishonest practices such as cheating and collusion [7]. Furthermore, the emergence of sophisticated deep learning (DL) tools like ChatGPT, while transformative in the educational applications, raises new concerns about their potential misuse in facilitating academic dishonesty.

The critical question that arises is, "Can current computing technologies effectively mitigate the risks of academic dishonesty in synchronous online assessments?" Extensive research and evidence indicate a rising prevalence of assessment misconduct, commonly termed e-cheating, in online educational settings [8], [9], [10], [11], [12], [13], [14]. Conventional misconduct detection methods, often designed as single-stage detection systems in nature, face challenges in identifying subtle forms of cheating and typically require extensive oversight, which diminishes their practicality [15], [16], [17].

These approaches are tended to focus on a single type of data or analytical method, limiting their ability to capture the complex patterns of e-cheating. Imagine a system designed only to spot plagiarism; it might do well in that specific task but could miss other cheating forms like students working together secretly (collusion) or using more clever, combined methods to cheat. The world of online learning is varied - students behave and respond differently, and their actions don't always follow a predictable pattern [18], [19]. This variety makes it hard for systems that only have one way of looking at things (single-stage detection systems) to accurately adapt and identify all the different ways students might misconduct. The dynamic nature of online learning, where student behaviours are not uniform, poses a significant challenge to the adaptability and accuracy of single-stage detection systems.

In response to these challenges, there is a growing need for multi-stage detection systems. We therefore present a comprehensive strategy that employs multi-stage strategy by implementing both behavioural and technological interventions, addressing the inherent limitations of previous methods. Our method aims to provide a nuanced understanding of the multifaceted nature of cheating behaviours, capturing a spectrum of dishonest activities that are often overlooked by singular analytical techniques.

Our study's novelty lies in the development of an intelligent agent that proactively monitors candidates' activities during synchronous online assessments. Specifically, our approach introduces an intelligent agent (multi-IA) comprises two key modules: the IP Detector, responsible for identifying suspicious behaviours through IP address monitoring, and the DL-based Behavioural Monitor, which analyses candidates' response behaviours. By unifying these two technological modules, our multi-IA system offers a more holistic and robust framework for detecting e-cheating. This multi-stage strategy not only enhances the detection of academic dishonesty but also reinforces the overall credibility and integrity of synchronous online assessments. This is especially crucial in an era where sophisticated tools like ChatGPT could unintentionally facilitate academic dishonesty by providing instant, complex responses. Through this strategy, we strive to equip educators with a comprehensive and efficient solution that curbs the risks of e-cheating and fortifies the credibility of synchronous online assessments.

Hence, the main contributions of this paper are outlined as follows:

- Advancing the field of academic integrity by proposing a multi-stage detection system, which synergises geographical verification and behavioural analysis. This strategy represents a significant step forward in employing computational technologies to combat academic dishonesty in online learning environments, addressing a gap in current research.
- Introducing an innovative e-cheating multi-IA for advanced misconduct detection. The multi-IA amalgamates an IP Detector and a DL-based Behavioural Monitor named DenseLSTM, which, to the best of our knowledge, represents a novel combination of techniques not explored in the existing literature.
- Conducting an exhaustive analysis of IP address detection and behavioural patterns monitoring in synchronous online assessments, enhancing understanding of academic integrity challenges in such contexts. This analysis provides valuable insights into the effectiveness of multi-stage detection in identifying e-cheating.

The remainder of this paper is organised as follows: Section II reviews the related works of existing approaches. Section III describes the methodologies employed by the multi-IA, while Section IV details the data collection process and its procedures of this study. Section V presents the experimental setup, followed by the analysis of results and discussion. A conclusion is summarised in the final section.

II. RELATED WORK

This section focuses on reviewing relevant works that explore the use of network technologies and behaviour analysis as effective approaches to detect and identify "suspected" cases of misconduct during synchronous online assessments, as tabulated in Table 1.

To provide a comprehensive comparison of our proposed multi-IA system with existing methods for detecting online assessment misconduct, we have conducted a thorough analysis of various network technology-based and behaviour analysis-based approaches. In the realm of network technology-based methods, studies such as [22] and [23] have explored the use of wireless communication and dynamic profile questions to identify and prevent cheating. While these approaches have shown promise in detecting certain forms of misconduct, they may face challenges in identifying more

TABLE 1. Summary of related works.

Reference	Assessment	Method	Highlights
[20], [21], [22]	MCQ	Network Tech.	- Framework for secure online exams with wireless technology. - Implement IP address for candidate identification.
[23]	N/S	Network Tech.	- Implement dynamic profile questions to prevent misconduct. - Mitigate exam impersonation through dynamic profile questions.
[24]	N/S	N/S	- Secure online exams with randomised questions and tab locking.
[25], [26]	N/S	Network Tech.	- Address the challenges of accessibility and reliability of network connection. - Emphasise prevention and ongoing research for integrity.
[27], [28]	N/S	Behaviour Anal.	- Implement a proctoring system to monitor candidates' behaviour.
[29], [30]	N/S	Behaviour Anal.	- Evaluate response patterns and monitor candidate behaviour. - Emphasise the similarity of response patterns among different candidates.
[31], [32], [33], [34]	N/S	Behaviour Anal.	- Implement face detection to monitor candidate behaviour.

MCQ = Multi-choice question; *N/S* = not specified;

subtle or complex cheating behaviours. Our multi-IA system addresses these limitations by incorporating both network analysis and behavioural monitoring, enabling a more comprehensive detection of various types of misconduct.

Behaviour analysis-based methods, such as those proposed by [29] and [30], have focused on evaluating response patterns and monitoring candidate behaviour to identify cheating. These approaches have demonstrated the importance of considering factors such as response consistency and completion times. However, they often rely on manual analysis by invigilators, which can be subjective and challenging to scale for large-scale assessments. Our multi-IA system overcomes these limitations by automating the behavioural analysis process using DL techniques, enhancing the objectivity and scalability of the detection mechanism. Furthermore, our proposed system distinguishes itself from existing methods by integrating both network analysis and behavioural monitoring into a unified framework. By leveraging the strengths of both approaches, our multi-IA system offers a more robust and comprehensive solution for detecting online assessment misconduct. The combination of IP address monitoring and deep learning-based behavioural analysis enables our system to identify a wider range of cheating behaviours while minimising false positives and false negatives.

In summary, our multi-IA system advances the state-of-the-art in detecting online assessment misconduct by addressing the limitations of existing network technology-based and behaviour analysis-based methods. By providing a more comprehensive and automated approach, our system contributes to the development of effective strategies for ensuring the integrity of online assessments in educational settings.

A. NETWORK TECHNOLOGY

The rapid growth of network technology has revolutionised remote access to digital resources, facilitating continuous monitoring of candidates' device usage and network traffic. This enables the identification of suspicious activities and potential instances of academic dishonesty by leveraging wireless communication technology. [22] conducted a study that highlighted the significance of monitoring candidates' use of internet messengers for communication, as most assessment misconduct incidents occurred due to unmonitored communication through these platforms. It is important to note that even without internet messengers, candidates can still engage in discussions if they are in the same physical location.

A study conducted by [23] sheds light on the mechanisms of wireless communication in synchronous online assessments. The study implemented a system that monitored participants during an assessment by employing dynamic profile questions within an online course. The results provided valuable insights, indicating that participants who engaged in cheating through impersonation showed a higher tendency to share information using mobile devices, resulting in distinct differences in their response times compared to non-cheating participants. In addition, [24] suggested the randomisation of question banks as a method to enhance the integrity of the assessment process. While randomising the question sequence can provide additional measures for detecting and addressing potential instances of misconduct, it is important to note that anomalies or patterns of unusual behaviour can still be identified and investigated, even when the question sequence is randomised.

A further review conducted by [20] also emphasised the significance of strengthening network protocols to enable techniques such as ''suspected'' case detection to address synchronous online assessment misconduct. Candidates were remotely monitored during an assessment, and their behaviours were simultaneously assessed with randomised questions presented through a dynamic profile. Results indicated that candidates who impersonated shared the most information through instant messaging, and consequently, their response times were significantly different from other candidates. Additionally, [21] highlighted a

potential challenge associated with the use of IP addresses for candidate identification in the context of addressing assessment misconduct. They point out that relying solely on IP addresses may create confusion, especially in scenarios where multiple candidates are located in the same building. This limitation emphasises the need for additional techniques to enhance the prevention and detection of assessment misconduct during synchronous online assessments.

Moreover, the accessibility and reliability of internet connections pose significant challenges in conducting synchronous online assessments. For instance, [25] and [26] have shown that not all candidates have access to reliable internet connections in their homes, especially in remote or rural areas. This lack of access can impede their participation in synchronous online assessments and limit their ability to fully engage in the learning process. Even for candidates with internet access, the speed and stability of the connection may not always meet the requirements for smooth online experiences. Fluctuations in internet speed and frequent disruptions can interrupt the assessment process, causing frustration and potential disadvantages for certain candidates.

B. BEHAVIOUR ANALYSIS

Behaviour analysis has received significant attention as a robust method to identify potential instances of cheating and plagiarism, with strong support from statistical evidence [27], [28]. However, to design effective behaviour analysis techniques, it is essential to understand the psychological and sociological theories that underpin human behaviour. For instance, social cognitive theory emphasises the role of observational learning and social influences in shaping behaviour, providing valuable insights into how candidates may engage in cheating or academic dishonesty. Moreover, considering the contextual factors that influence behaviour, such as the competitive nature of the assessment environment or the presence of peer pressure, enhances the accuracy and reliability of behaviour analysis techniques.

Various approaches have been proposed to leverage behaviour analysis in effectively detecting assessment misconduct. Early work done by [35] and [36] suggested the use of proctoring security software, which allows invigilators to monitor candidates' computer screens through an instructor control view. While these approaches have shown promise, they rely on the installation of specialised software and may face challenges in cases where the software is not installed or when there is an unstable connection. Notably, [29] emphasised the importance of evaluating the regularity of candidates' response patterns, including factors such as the regression of testing time on response scores or the similarity of response patterns among different candidates.

Furthermore, [30] sheds light on the significance of monitoring candidate behaviour to identify cheating, specifically through the analysis of the number of revisions made and the pattern of response times. By scrutinising these behavioural indicators, it becomes possible to detect irregularities that may indicate dishonest practices. Although the response pattern is considered of lesser importance in this study, it still contributes to the overall detection process by providing additional information and potential evidence of cheating.

Additionally, several studies conducted by [31], [32], [33], and [34] have explored the application of face detection, specifically by monitoring the yaw angle among candidates using cameras, to detect potential instances of cheating or suspicious behaviour during assessments. These biometric approaches offer valuable insights into enhancing assessment integrity. However, a significant challenge arises from the reliance on invigilators to monitor candidates during the assessment and ensure compliance with rules and regulations. Requiring candidates to attach cameras for monitoring purposes can introduce logistical complexities and may not be feasible or suitable for all assessment scenarios, thereby limiting the widespread implementation and effectiveness of these biometric methods.

However, despite the effectiveness of behaviour analysis in detecting misconduct, a significant limitation of these approaches lies in their reliance on manual analysis by invigilators. This reliance introduces subjectivity, potential human error and practical challenges in scaling the detection process for large-scale assessments. Therefore, further research and development of automated systems or intelligent algorithms are necessary to enhance the objectivity, accuracy and scalability of behaviour analysis in combating e-cheating.

C. MOTIVATION

In the realm of countering fraudulent activities, various methods have been explored, including the use of network technology and behaviour analysis, as summarised in Table 1. While each approach has its merits, they also have its limitations in effectively preventing e-cheating. Additionally, many institutions face financial constraints that hinder their ability to invest in expensive anti-cheating technologies such as online proctoring or secure browsers. Consequently, there is a significant demand for a comprehensive and intelligent solution that combines these approaches, thereby improving the detection and prevention of cheating behaviours in synchronous online assessment.

In this study, we propose a new case study focused on detecting abnormal behaviour of candidates during synchronous online assessments and, consequently, preventing e-cheating. Our approach involves the utilisation of a multi-IA functioning as a proactive monitoring system. The multi-IA is designed with advanced network protocol detection and DL techniques, enabling it to actively monitor candidates' activities during synchronous online assessments. By analysing behaviour patterns, response times and other relevant factors, the multi-IA can identify abnormal behaviour indicative of potential cheating.

III. METHODOLOGY

To enhance the clarity of our system, we present the architectural framework of a multi-IA system meticulously designed

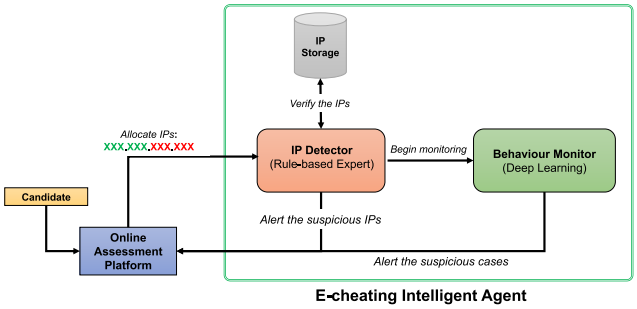


FIGURE 1. Illustration of the E-Cheating Multi-Intelligent Agent (multi-IA) framework. The framework comprises two key modules designed to detect and prevent misconduct in synchronous online assessments. The modules include the IP Detector, responsible for identifying "suspicious" behaviours through IP address monitoring, and the DL-based Behavioural Monitor, which analyses candidates' response behaviours.

for identifying and preventing suspicious behaviours during synchronous online assessments. The foundational components of our multi-IA system consist of two key modules: a network IP Detector and a Behavioural Monitor. These modules work in tandem to provide a comprehensive approach to misconduct detection. The intricate details of the system structure are elucidated in Figure 1, offering a visual representation of its functioning and the dynamic interaction between the IP Detector and Behavioural Monitor.

Our multi-IA system is strategically developed to align with the intricacies of online assessments prevalent in colleges and universities. It accommodates a diverse range of assessment formats, including multiple-choice questions (MCQ), True/False (T/F) questions and short-answer questions. Notably, our system's flexibility extends to its compatibility with any online assessment platform, making it a universal solution applicable across diverse infrastructures. This adaptability ensures seamless integration, regardless of the specific technology or Learning Management System (LMS) in use. Moreover, the multi-IA system is designed to scale effectively, meeting the requirements of different class sizes. This scalability feature highlights the system's robustness and its ability to perform optimally in various educational settings.

A. IP DETECTOR

Referring to Figure 1, we present the IP Detector module, a crucial component within our framework, designed to bolster first-stage authentication through IP address monitoring. IP addresses, expressed as a set of four numbers (e.g., 192.158.1.38), serve a crucial role in identifying and tracking network connections, validating the legitimacy of a candidate's location, and uncovering suspicious activities related to unauthorised access. Additionally, IP addresses offer insights into the consistency of a candidate's online behaviour, enabling the detection of anomalies or discrepancies across assessment sessions, indicative of potential misconduct or unauthorised behaviour.

Algorithm 1 IP Detector

INPUT: IP address from a new candidate E
OUTPUT: Classification result D

Let IP_{DB} be the list of real-time IPs of size N
if E is not in IP_{DB} **then**
  Add new IP address E to the database IP_{DB}
  return an assigned random question set D to E
else
  for $i \rightarrow 1$ to N **do**
    if E is found in $\text{IP}_{\text{DB}}[i]$ or E has high similarity in $\text{IP}_{\text{DB}}[i]$ **then**
      return a different question set D to E
    end if
  end for
end if

In our framework, we utilise the similarity between the initial and last two numbers of the IP address as a triggering factor for detecting suspicious cases. Addressing concerns about the reliance on IP consistency, our methodology considers the prevalence of static IP addresses allocated by many internet service providers to residential customers. As corroborated by various studies [37], [38] and confirmed through our analysis in Section V-C1, static IPs exhibit infrequent changes, establishing them as a reliable metric for our monitoring efforts. By incorporating this awareness into our analysis, we aim to ensure the robustness and applicability of our monitoring efforts across a spectrum of educational settings.

Empowering the IP Detector module to effectively track and monitor candidate IP addresses, our approach contributes to the identification of suspicious behaviours, thereby augmenting the overall security of synchronous online assessments. As depicted in Figure 1, the framework seamlessly retrieves the IP address of each candidate. Every candidate is initially assigned a random assessment question set, such as Set A. Upon candidate login, the IP Detector module retrieves their IP address from the IP storage. If the IP address is flagged as "suspicious" in the IP list (as indicated by the "red" numbers in Figure 1), the multi-IA system promptly notifies the invigilator and recommends generating a different set of questions, such as Set B, for the candidate.

Algorithm 1 succinctly outlines the complete IP detection process. The impacts of this module are further explored in Section V-A, providing a comprehensive evaluation across our online assessment formats and addressing concerns raised regarding the efficacy and applicability of our method.

B. BEHAVIOURAL MONITOR

Assuming the successful assignment of assessment question sets by the IP Detector, as illustrated in Figure 1, the Behavioural Monitor module plays a crucial role within the multi-IA framework. This module is intricately designed

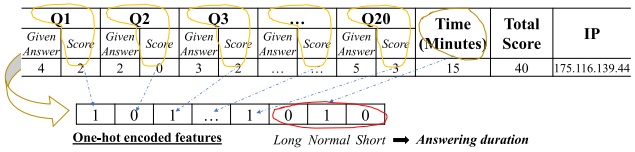


FIGURE 2. Generation of the one-hot encoded feature. The encoded feature vector is created by representing the scores of each question and the time taken for completion. Each question score is represented as a binary value, where a value of *1* indicates a correct answer and a value of *0* represents an incorrect answer. The time taken for completion is also included as a feature in the vector.

to identify and flag abnormal behaviours in candidates' answering patterns and response times. To accomplish this, we have implemented this module in our multi-IA system utilising a DL approach, namely DenseLSTM. This module analyses the candidates' behaviour by examining their sequence of responses, the time taken to answer questions and performance disparity. When the multi-IA system detects abnormal behaviour, it promptly triggers an alert to the invigilators, enabling them to investigate the flagged candidates and take appropriate actions.

1) PREPROCESSING

A crucial preprocessing step was conducted to represent raw data records for the development of our multi-IA system. Each raw data record was transformed into a one-hot encoded feature, which effectively captured the behaviour of the candidate. Specifically, the raw data record included information such as the pattern of answer results for the N questions, the total time (in minutes) taken to complete the assessment and the final score. This preprocessing step aligns with the methodology advocated by [23] and [29], which underscored the significance of scrutinising the consistency within candidates' response patterns. By encoding these aspects of behaviour into a format amenable to DL, the multi-IA system gains the capability to subtle patterns and correlations within the data—nuances that may elude human evaluators.

Specifically, we define an encoded input vector as $R \in 1 \times M$, where $M = 23$. The first 20 elements represent the candidate's responses to the N questions, where N is set to 20. These elements contain $[1, 1, 0, \cdots, 0]$ with the values 1 and 0 denoting a correct or incorrect response, respectively. The last three elements of M define the answering duration, e.g., long, normal or short. The candidates' behaviours are classified into either "normal" or "suspected". To identify these behaviours, we followed the concept by [23] to analyse behaviours, which will be discussed in Section IV-A. Here, two factors were considered when defining the speed of answering: the number of questions answered and the difficulty level for each question. Figure 2 summarises the data representation, showing the processing of raw data into one-hot encoded features.

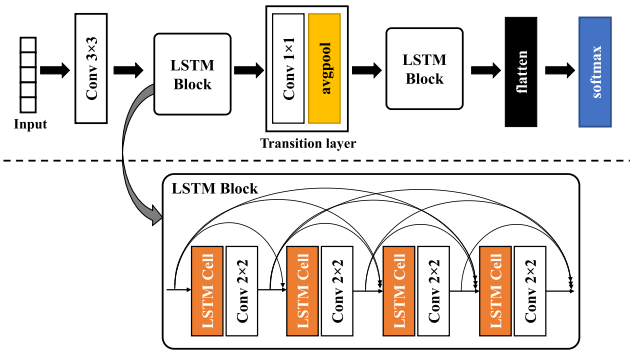


FIGURE 3. The architecture of the DenseLSTM. The network consists of two LSTM blocks, enabling the extraction of enhanced representations for understanding behaviour patterns related to e-cheating.

2) NETWORK STRUCTURE

Utilising a combination of the Long Short-Term Memory (LSTM) network [39] and the densely-connected approach [40], the multi-IA incorporates a new DL network called DenseLSTM, as illustrated in Figure 3. This network is designed to analyse and monitor candidates' behaviours, distinguishing between "normal" and "suspected" patterns. It is represented as an input vector R, which undergoes processing within DenseLSTM. To extract relevant features from R, it is passed through a *conv* layer, followed by two LSTM blocks denoted as B with concatenation operator H. Each x layer within B contains an LSTM cell, a *conv* layer with a filter size of 2×2, a rectified linear unit (ReLU) activation function and dropout layers, which are repeated by l times. By adopting the proposed LSTM block, the network can extract better feature representations, strengthening its feature activation for predicting potential e-cheating. The structure of B is expressed as follows:

$$B = H_l([x_0, x_1, x_2, \cdots, x_{l-1}]), \tag{1}$$

where x_0 to x_{l-1} denotes feature outputs. We set $l = 4$ in the first LSTM block and $l = 8$ in the second LSTM block. Furthermore, a transition layer is exclusively applied in the first LSTM block, incorporating *conv* 1×1 and *avgpool* operations. The *conv* 1×1 operation employs a filter size of 1×1 within the *conv* layer. The purpose of this transition layer is to manage the complexity of the feature maps.

In the training stage, we utilise softmax cross-entropy $\mathcal{L}$ to compute the loss between the logit vector $\mathbf{Y}$ and the corresponding encoded label, as shown in Equations (2) and (3):

$$\mathcal{L}(Y) = -\sum_{i}^{E}\sum_{j}^{C} L_{ij} \log(\text{softmax}(Y)_{ij}), \tag{2}$$

$$\text{softmax}(Y)_{ij} = \frac{\exp(Y_{ij})}{\sum_{j}^{C} \exp(Y_{ij})}, \tag{3}$$

where L, E and C denote class labels, the number of training samples in $\mathbf{Y}$ and the number of classes, respectively.

IV. DATASET

Considering the unique nature of our study environment, where no existing works have been conducted, we have carefully designed our procedures for detecting synchronous online assessment misconduct at an anonymous university, as outlined in Section III. Consequently, the specific details about the institutional context, such as the exact time frame and demographic specifics, have been omitted to comply with these regulations.

A. QUESTIONNAIRE STRUCTURE

We acknowledge that various environmental factors within each institution, such as institutional policies, LMS features and available technical infrastructure, play a significant role in this process. Additionally, it is important to address the challenges posed by environments where additional device attachments are not available or weak connections exist. We seek to gain insights into the nature of misconduct and explore potential strategies to effectively address them within such constrained settings.

The question bank consisted of approximately 60 questions that were carefully crafted to assess various aspects of the candidates' knowledge and skills in the field of AI. These questions were categorised into different types, including MCQ, T/F and short-answer questions. During each assessment, a set of 20 questions is selected, covering three levels of difficulty: *easy*, *moderate*, and *advanced*.

To accommodate the variation in learning performance across different semesters, the number of questions in each assessment is adjusted accordingly. For the ''mid-term'' assessment, the general distribution includes 7–9 *easy* questions, 6–8 *moderate* questions and 2–4 *advanced* questions. On the other hand, the ''final-term'' assessment typically consists of 1–3 *easy* questions, 10–14 *moderate* questions and 2–4 *advanced* questions. Specifically, for the *easy* and *moderate* difficulty levels, which are typically associated with MCQ and T/F questions, we focused on these question types. Additionally, for the *advanced* level, we incorporated short-answer questions to provide a more challenging assessment. However, the specific allocation of questions within each difficulty category can be customised based on the instructor's settings in the LMS. This flexibility allows for adaptability and tailoring of assessments to suit the unique characteristics and abilities of the candidates in each semester.

B. IP ADDRESS RECORD

It is important to note that the LMS at an anonymous university has certain limitations. While it allows us to record the candidate's scores on each question, along with their response answer, total response time, total grades and IP address, there may be additional features or data points that are not captured by the system.

Nevertheless, we hypothesise that by analysing these available features, we can still identify suspicious behaviour, such as unusually short or long response times, multiple logins from different devices, or deviations from baseline behaviour patterns. Further analysis of these features is crucial in maintaining assessment integrity and ensuring a fair evaluation of candidates' knowledge and skills, which will be discussed in detail in Section V-C2.

C. DATA COLLECTION AND PROTOCOLS

1) TRAINING SET

For training, we curated a comprehensive and diverse learning dataset. This dataset includes 192 records representing a collection of candidate behaviours from recent assessment periods. The records encapsulate two separate sets of candidate assessments, each contributing to a broad range of behaviours indicative of potential misconduct and IP address during online examinations.

Our objective in creating this dataset was to capture a wide spectrum of behaviours and IP address that could potentially be interpreted as misconduct during synchronous online assessments. The dataset incorporates both observed scenarios specifically designed to simulate potential misconduct and real data reflecting actual candidates' behaviour in answering assessments. These observed scenarios aim to replicate various types of misconduct that can occur during synchronous online assessments, such as unauthorised collaboration, cheating, or using advanced resources such as ChatGPT. By combining these two types of data, we aimed to provide a more comprehensive understanding of potential misconduct behaviours and improve the training of our AI system.

2) TESTING SET

To evaluate the performance of our system and gain insights into candidates' behaviours and performance, we collected a comprehensive evaluation dataset. This dataset is comprised of 569 records, collating the responses of 301 candidates across three semesters: *Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*. The selected periods span an extensive time frame, including a mix of assessments that typically occur throughout an academic cycle. This strategic selection ensures that the dataset reflects a wide-ranging set of conditions, candidate behaviours and IP address, providing a solid foundation for comprehensive performance evaluation of the multi-IA system.

The evaluation dataset provides a holistic view of the candidates' performance throughout the academic year. The candidate records were meticulously compiled to capture essential information such as response patterns and network traffic data. Through the analysis of this diverse evaluation dataset, our study aims to offer a comprehensive analysis of candidates' behaviours and performance in synchronous online assessments, further enhancing the effectiveness of our AI system.

D. ETHICAL AND ASPECTS

In conducting this study within the realm of educational technology and assessment integrity, meticulous attention has

been given to privacy and data protection considerations. In adherence to stringent data protection regulations, the collection and analysis of data, encompassing candidates' IP addresses and behavioural patterns, are conducted with utmost care. Measures are implemented to ensure the confidentiality and security of collected data, preserving the anonymity of candidates and safeguarding their personal information.

V. EXPERIMENT

This phase of the study focuses on a comprehensive exploration of the multi-IA system's stability, delving into intricate technical aspects. The evaluation encompasses an in-depth analysis of the performance and stability of the IP Detector module, followed by an assessment of the reliability of the behavioural monitor within our multi-IA system. In addition, we also added ablation study to support certain concerns that may be address in our online assessment settings.

A. EVALUATION ON IP DETECTOR

The evaluation of the IP Detector has yielded substantial insights into its effectiveness of detecting suspicious IP addresses and its role in identifying potential instances of misconduct. This assessment involved a thorough analysis of IP detection results across multiple final-term semesters, specifically *Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*. The IP Detector algorithm, structured on a rule-based case following established standards [37], [38], guided the geophysical visualisation of collected IP addresses during assessments, aiming to pinpoint suspicious cases.

In Figure 4, we meticulously present a nuanced visual analysis, distinguishing between ''suspected'' and ''normal'' instances across diverse semesters. The graph employs colours such as ''grey'', ''orange'' and ''red'' to represent ''individual'', ''neighbouring'' and ''exact'' candidate location, respectively, which offers a detailed examination of potential misconduct cases. The visual insights provide transparency into the IP Detector's performance, underscoring its ability to discern geographical locations effectively. These findings unequivocally underscore the capability of our IP Detector module, enhancing the multi-IA system's effectiveness in early-stage identification of potential misconduct. By systematically identifying IP addresses with unusual locations, the multi-IA system can strategically focus its investigation on potential instances of misconduct, thereby upholding the integrity of the synchronous online assessment process.

The detection of suspicious IP locations provided valuable insights for subsequent investigation. It enabled the multi-IA system to concentrate its efforts on specific IP addresses that required closer examination, streamlining the process of identifying potential instances of misconduct. By targeting IP addresses with unusual locations, the multi-IA system effectively safeguarded the integrity of the synchronous online assessment process and helped prevent misconduct from occurring.

B. PERFORMANCE COMPARISON ON BEHAVIOURAL MONITOR

1) EXPERIMENTAL SETTINGS

To objectively evaluate the performance of our multi-IA system in a practical setting, we conducted experiments over three semesters: *Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*. The detailed evaluation protocol of these experiments are thoroughly discussed in Section IV-C2. Our evaluation aimed to assess the performance of the multi-IA system by comparing it against the overall instructor's opinion and two widely used Deep Learning (DL) algorithms, namely Multi-layer Perceptron (MLP) [41] and LSTM. This evaluation allowed us to assess the effectiveness of our multi-IA and its performance in comparison to alternative methods.

During training, we adhered to the protocols outlined in Section IV-C1. The deep learning algorithms were configured with the learning rate to 1.0×10^{-5} and employ the AdamOptimizer [42], with weight decay and momentum values of 1.0×10^{-4} and 0.9, respectively. The batch size is set to 16, and the training process is conducted over 100 epochs.

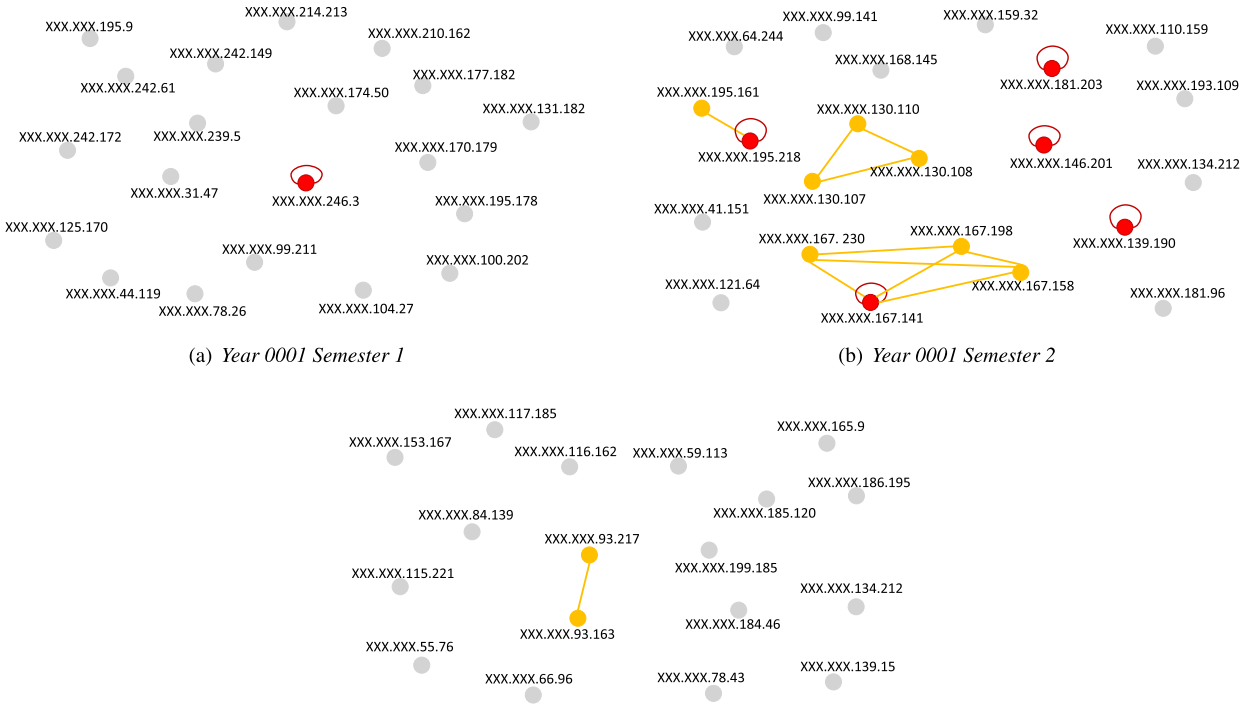
2) EVALUATION METHODS

Our evaluation method employs stringent metrics to assess the performance and reliability of our DL-based Behaviour Monitor. We utilise fundamental metrics such as accuracy, Receiver Operating Characteristic (ROC) curve, Area Under the Curve (AUC) and confusion matrix.

Specifically, accuracy is crucial as it measures the ratio of correctly predicted instances, providing a direct indicator of the model's precision. The ROC curve is instrumental in offering insights into the sensitivity-specificity trade-off, thereby elucidating the Behaviour Monitor's discriminatory power across different decision thresholds. AUC quantifies the overall performance and the model's ability to distinguish between classes, which serves as a comprehensive metric capturing the Behaviour Monitor's effectiveness across various aspects of assessment integrity. Additionally, the confusion matrix plays a critical role, helping identify specific areas where the model makes errors and guiding potential adjustments for enhanced performance.

3) EXPERIMENTAL RESULTS

Upon conducting an in-depth evaluation, our study's robust experimental results, presented in Table 2, substantiate the efficacy of our proposed Behavioural Monitor for synchronous online assessment misconduct detection. The DenseLSTM network, a key component of our multi-IA system, demonstrated exceptional accuracy, achieving an impressive overall accuracy rate of 87%. This remarkable outcome performance consistently surpassed other evaluated methods, including MLP, LSTM, and even human assessments represented by the instructor's opinions. Notably,



(a) *Year 0001 Semester 1*

(b) *Year 0001 Semester 2*

(c) *Year 0002 Semester 1*

FIGURE 4. Geospatial Mapping of IP addresses across distinct testing sets (*Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*): The IP Detector demonstrates proficient identification of "suspected" IP addresses, delineated by nodes colour-coded in "orange" and "red" on the graph. Notably, the presence of "red" nodes with connected edges signifies that candidates are located in the same place. This figure is best viewed digitally.

the DenseLSTM model showcased remarkable accuracy rates of 91.96% in *Year 0001 Semester 1*, 88.78% in *Year 0001 Semester 2* and 87% in *Year 0002 Semester 1*, establishing it as a promising solution for effective misconduct identification in synchronous online assessments.

To provide a nuanced assessment of our multi-IA system's discrimination capabilities, we employed the ROC curve and the AUC analyses. The results, illustrated in Figure 5 and summarised in Table 2, showcase the exceptional performance of the DenseLSTM network, surpassing alternative methods. The ROC curves vividly illustrate DenseLSTM's effectiveness in distinguishing "normal" and "suspected" behaviours during synchronous online assessments. Specifically, for the testing sets from *Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*, the DenseLSTM network exhibited impressive AUC values of 0.9561, 0.8997, and 0.8876, respectively. These remarkable AUC values also highlight the model's robust ability to accurately classify candidates' behaviours, ensuring a thorough evaluation of our multi-IA system's performance.

Further confusion matrix analysis unveiled an overall error rate ranging from 11% to 18%. False alarms, triggered by the multi-IA system, occurred when it incorrectly identified candidates for potential misconduct due to response time variations. Conversely, missed detections transpired when instances of genuine misconduct evaded detection by the system. These observations shed light on the challenges associated with accurately discerning authentic instances of misconduct and highlight factors contributing to suspicious behaviour.

C. ABLATION STUDY

1) ANALYSIS OF STATIC IP ADDRESS

This study examines the presence of statically assigned IP addresses and assess the stability of IP addresses during the assessment period. Table 3 displays sample records of IP addresses for candidates across different semesters. It is observed that a significant portion of candidates' IP addresses remained static or unchanged during the assessment period.

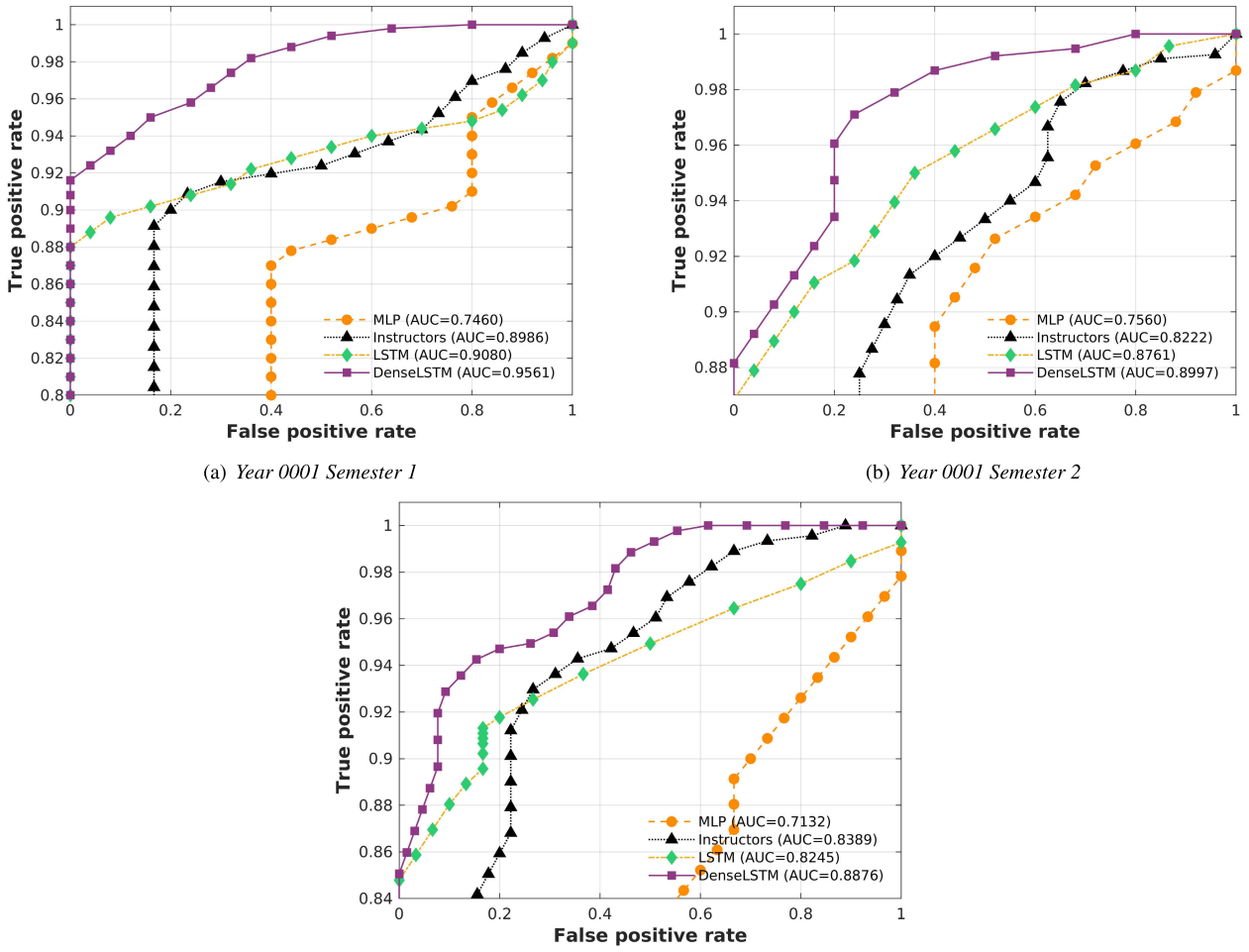
As shown in Table 3, most of the IP addresses remained static during the mid-term and final-term periods in each semester. This implies that the candidates' devices were consistently connected to the same network or using the same internet service provider, resulting in consistent geographical locations associated with their IP addresses. Thus, the static nature of the IP addresses becomes an important factor in establishing a baseline reference for each candidate's expected IP address behaviour during the assessments.

Another important aspect to consider is the use of virtual private networks (VPN) that can potentially bypass the IP Detector. VPN services can be employed by candidates to obfuscate their authentic IP addresses, thus misleading

TABLE 2. Performance evaluation of synchronous online assessments for the *Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*. Figures of highest accuracy, highest AUC value and lowest error rate are highlighted in bold.

Method	*Year 0001 Semester 1*		*Year 0001 Semester 2*		*Year 0002 Semester 1*		Overall Confusion Matrix			
	Accuracy	*AUC*	*Accuracy*	*AUC*	*Accuracy*	*AUC*	*TP*	*TN*	*FP*	*FN*
MLP	72.77%	0.7460	71.43%	0.7560	68.00%	0.7132	74.69%	50.07%	46.23%	25.31%
LSTM	89.29%	0.9080	85.71%	0.8761	79.00%	0.8245	84.98%	58.89%	41.11%	15.02%
DenseLSTM	**91.96%**	**0.9561**	**88.78%**	**0.8997**	**87.00%**	**0.8876**	**88.05%**	**86.11%**	**17.86%**	**11.95%**
Instructors	86.35%	0.8986	81.63%	0.8222	75.00%	0.8389	83.62%	58.23%	41.77%	16.38%

TP=True Positive; *TN*=True Negative; *FP*=False Positive; *FN*=False Negative



(a) *Year 0001 Semester 1*

(b) *Year 0001 Semester 2*

(c) *Year 0002 Semester 1*

FIGURE 5. Performance comparisons on final terms in (a) *Year 0001 Semester 1*, (b) *Year 0001 Semester 2* and (c) *Year 0002 Semester 1*. This figure is best viewed digitally.

the monitoring system. However, the Behavioural Monitor module in multi-IA can still detect suspicious behaviour, even if the IP address is masked by a VPN. This additional layer of analysis helps to overcome the limitations posed by VPN usage and ensures a more comprehensive assessment of candidate behaviour. The details of this Behavioural Monitor module will be discussed in later sections.

2) ANALYSIS OF RESPONSE PATTERN

We followed the behavioural classification concept proposed by [23] to categorise the recorded response patterns as either "normal" or "suspected". This allows us to better understand their behaviour patterns and detect any irregularities or suspicious patterns that may indicate misconduct. As mentioned in Section IV-A, the assessment consisted of 20 questions and the levels of difficulty is depending on each semester.

TABLE 3. Sample records showing static IP addresses from *Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*. Figures of identical candidates with different IP addresses are highlighted in bold.

	Year 0001 Semester 1			*Year 0001 Semester 2*			*Year 0002 Semester 1*	
#	IP Addresses		#	IP Addresses		#	IP Addresses	
	Mid-term	Final-term		Mid-term	Final-term		Mid-term	Final-term
1	175.xxx.xxx.44	175.xxx.xxx.44	1	211.xxx.xxx.75	211.xxx.xxx.75	1	180.xxx.xxx.98	180.xxx.xxx.98
2	211.xxx.xxx.62	211.xxx.xxx.62	2	222.xxx.xxx.212	222.xxx.xxx.212	2	124.xxx.xxx.111	124.xxx.xxx.111
3	125.xxx.xxx.50	125.xxx.xxx.50	3	116.xxx.xxx.139	116.xxx.xxx.139	3	58.xxx.xxx.149	58.xxx.xxx.149
4	58.xxx.xxx.182	58.xxx.xxx.18	4	1.xxx.xxx.73	(absent)	4	**203.xxx.xxx.136**	**61.xxx.xxx.228**
5	211.xxx.xxx.3	211.xxx.xxx.3	5	**222.xxx.xxx.158**	**121.xxx.xxx.40**	5	118.xxx.xxx.20	118.xxx.xxx.20
6	211.xxx.xxx.3	211.xxx.xxx.3	6	45.xxx.xxx.158	45.xxx.xxx.158	6	211.xxx.xxx.217	211.xxx.xxx.217
7	**61.xxx.xxx.62**	**121.xxx.xxx.164**	7	45.xxx.xxx.158	45.xxx.xxx.158	7	1.xxx.xxx.228	1.xxx.xxx.228
8	**182.xxx.xxx.51**	**112.xxx.xxx.188**	8	124.xxx.xxx.218	124.xxx.xxx.218	8	1.xxx.xxx.160	1.xxx.xxx.160
9	182.xxx.xxx.184	182.xxx.xxx.184	9	125.xxx.xxx.240	125.xxx.xxx.240	9	124.xxx.xxx.38	124.xxx.xxx.38
10	118.xxx.xxx.52	(absent)	10	118.xxx.xxx.112	118.xxx.xxx.112	10	124.xxx.xxx.117	124.xxx.xxx.117

Figure 6 illustrates the candidates' scores in several assessments: training set and testing set during mid-term *Year 0001 Semester 1*, mid-term *Year 0001 Semester 2* and mid-term *Year 0002 Semester 2*. Based on the analysis of the training set in Figure 6(a), which includes grades and response time, we observed that the majority of candidates exhibited "normal" behaviours and completed the assessments within a time range of 15 to 25 minutes. Specifically, we found that many candidates answered *easy* questions within 20 to 25 seconds, while *moderate* and *advanced* questions required approximately 40 to 50 seconds and 2 to 3 minutes, respectively. These findings provide evidence that aligns with the expected behaviour patterns and shed light on the typical time taken by candidates to complete different question types.

Moreover, the results for other semesters consistently indicated that several candidates conducted the assessments with "suspected" behaviours, highlighting potential irregularities in their responses. As depicted in Figures 6(b)–6(d), approximately 25–30% of the total candidates in each semester were flagged as "suspected" cases. These flags were raised due to various reasons, such as candidates completing the assessment significantly faster or slower than the average completion time and providing identical or near-identical responses to other candidates.

3) PERFORMANCE DISPARITIES

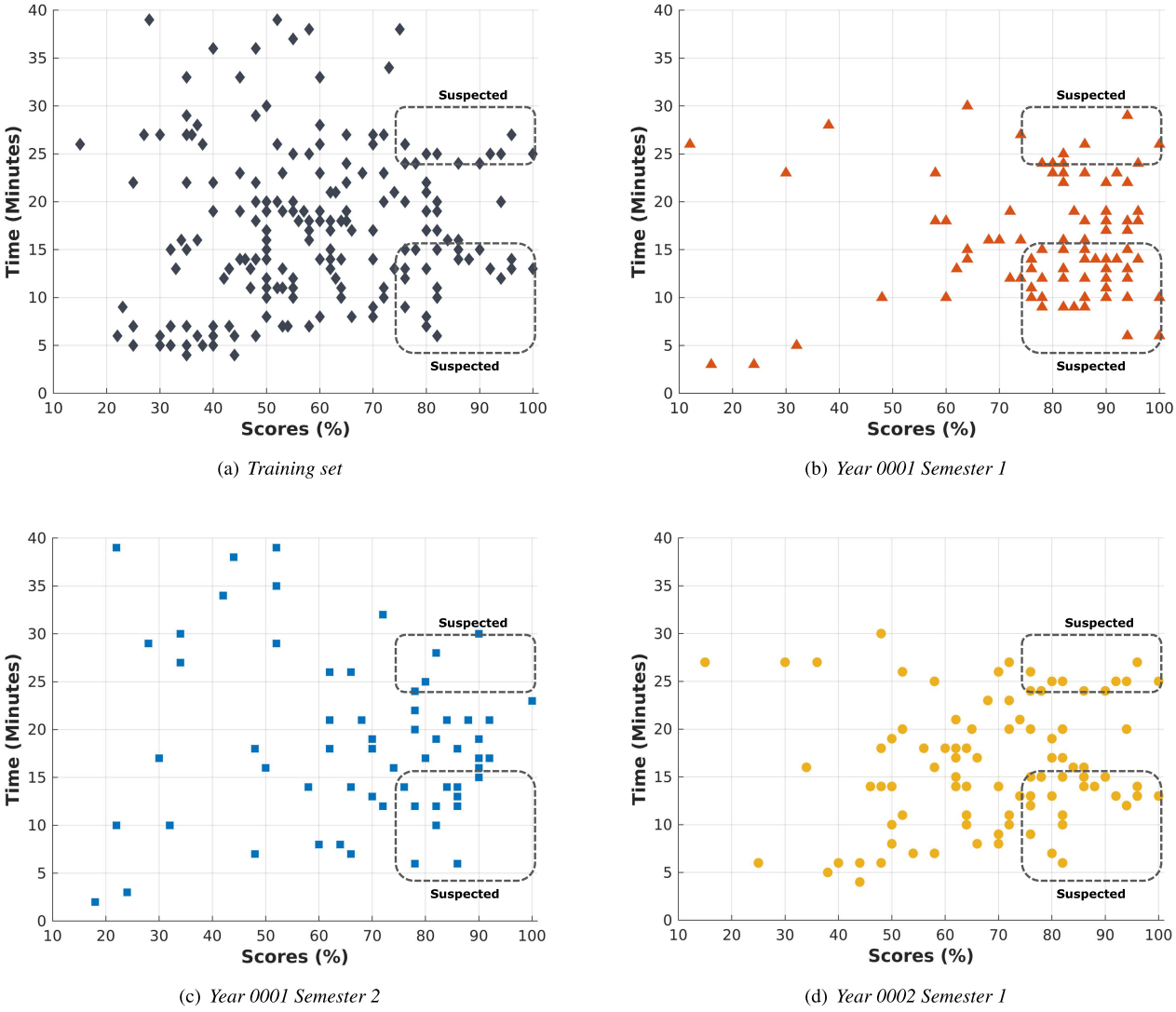
In addition, analysing performance disparities can also yield valuable insights into potential irregularities or misconduct. Here, the definition of performance disparity refers to significant differences in performance between different types of questions. Table 4 indicates that specific evidence has been observed, supporting the analysis. It is noted that certain candidates consistently achieve high scores on moderate and advanced questions but perform poorly on easier ones. However, it was noted that such behaviour often occurred when completing the assessments within 15 minutes or more than 30 minutes.

Further investigation into these cases reveals intriguing details. Upon closer examination of these cases, intriguing details emerge. Candidates displaying performance disparities often provide identical or nearly identical answers to other candidates. Such findings emphasise the importance of considering performance disparities, response consistency, and completion times when analysing candidate behaviours. The similarity in responses suggests a potential breach of academic integrity, such as unauthorised collaboration or the use of illicit resources. Moreover, their response times exhibit inconsistencies, with rapid completion of some questions followed by significantly longer duration for others. These observations emphasise the necessity fir meticulous scrutiny and thorough investigation to uphold the integrity and equity of the assessment process.

D. LIMITATION OF THE STUDY

Our study highlights several limitations and considerations that impact the scholarly depth and critical assessment of our approach:

- Reliance on IP Address for Candidate Identification: Our study acknowledges the limitation of relying solely on IP addresses for candidate identification, as it may create confusion, especially in scenarios where multiple candidates are located in the same building. This limitation emphasises the need for additional techniques to enhance the prevention and detection of assessment misconduct during online assessments.
- Connectivity Challenges: Our study recognises the challenges posed by the accessibility and reliability of internet connections, especially for candidates in remote or rural areas. It emphasises the need to address these connectivity challenges to ensure equitable access to reliable internet connections and promote fairness in online assessments.



(a) *Training set* (b) *Year 0001 Semester 1*

(c) *Year 0001 Semester 2* (d) *Year 0002 Semester 1*

FIGURE 6. **Visualisation showing the labels of "suspected" cases based on the total scores and completion time from (a) training set, (b) Mid-term *Year 0001 Semester 1*, (c) Mid-term *Year 0001 Semester 2* and (d) Mid-term *Year 0002 Semester 1*. The "suspected" cases are highlighted within the dotted boxes. This figure is best viewed digitally.**

TABLE 4. **Performance disparities in answering *easy*, *moderate*, and *advanced* questions across the mid-term assessments from *Year 0001 Semester 1*, *Year 0001 Semester 2* and *Year 0002 Semester 1*. This table presents several cases of candidates' performance in each semester, specifically highlighting the number of questions answered correctly for different difficulty levels.**

Semester	Candidate	Easy Q.	Moderate Q.	Advanced Q.	Scores	Time in Mins
Year 0001 Semester 1	XXX021	3/9	8/8	2/3	B+	16
	XXX060	3/9	8/8	2/3	B+	14
	XXX074	4/9	8/8	3/3	A	12
Year 0001 Semester 2	YYY006	4/8	8/9	2/3	A-	18
	YYY019	2/8	9/9	2/3	B+	20
	YYY058	2/8	6/9	2/3	B-	51
Year 0002 Semester 1	ZZZ082	6/12	5/5	3/3	A	13

- Behaviour Analysis Limitations: While our study discusses the application of behaviour analysis to detect assessment misconduct, it also acknowledges the challenges associated with behaviour analysis, such as the reliance on IP consistency and the potential limitations in environments where additional device

attachments are not available or weak connections exist, and the limited experimental dataset that mainly focused on MCQ, T/F and short-answer questions.

By transparently outlining these limitations, our study lays the groundwork for future research endeavours to refine and expand upon our methodologies, fostering a more comprehensive and robust approach to tackling the complexities of online assessment integrity.

VI. CONCLUSION

This study introduces an innovative approach to tackle the challenges of academic integrity in online assessments by presenting the development of a multi-IA system. Integrating an IP Detector with a Behavioural Monitor, the multi-IA system exhibits promising results in identifying potential instances of e-cheating and assessment misconduct. The empirical analysis reveals a significant percentage of candidates displaying ''suspected'' behaviours, highlighting the prevalence of irregularities in online assessment responses. The study underscores the importance of analysing performance disparities, response consistency, and completion times to identify potential breaches of academic integrity.

The novel use of IP Detector and Behavioural Monitor, constitutes a substantial contribution to the field. The system's capability to differentiate between normal and potentially fraudulent responses underscores its potential as a valuable asset for educational institutions and educators in preserving the integrity of online assessments. Additionally, the study emphasises ethical and privacy considerations associated with implementing such technology, ensuring responsible and equitable use in educational settings.

The findings of this study carry implications for ongoing technology development to mitigate the risks of academic dishonesty in online learning environments. By providing a more trustworthy and reliable educational environment, the proposed multi-IA system aims to empower educators with an effective and scalable solution to detect and address suspicious activities. Continued refinement and evaluation of integrated solutions, such as the multi-IA system, are imperative to ensure the credibility and fairness of online assessments amid evolving technological landscapes.

It is important to acknowledge the limitations of our study, such as the limited experimental dataset that mainly focused on MCQ, T/F and short-answer questions, as well as standard/higher bandwidth mobile networks. Future investigations should aim to augment the system's competencies by testing it across a variety of courses and in unconstrained environments, considering alternative delivery formats such as essays and addressing environmental factors like low bandwidth networks.

REFERENCES

[1] O. B. Adedoyin and E. Soykan, ''COVID-19 pandemic and online learning: The challenges and opportunities,'' *Interact. Learn. Environ.*, vol. 31, no. 2, pp. 863–875, Feb. 2023.

[2] M. A. Almaiah, A. Al-Khasawneh, and A. Althunibat, ''Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic,'' *Educ. Inf. Technol.*, vol. 25, no. 6, pp. 5261–5280, Nov. 2020.

[3] R. Ryu, S. Yeom, D. Herbert, and J. Dermoudy, ''A comprehensive survey of context-aware continuous implicit authentication in online learning environments,'' *IEEE Access*, vol. 11, pp. 24561–24573, 2023.

[4] B. Keresztury and L. Cser, ''New cheating methods in the electronic teaching era,'' *Proc. Social Behav. Sci.*, vol. 93, pp. 1516–1520, Oct. 2013.

[5] K. Jalali and F. Noorbehbahani, ''An automatic method for cheating detection in online exams by processing the student's webcam images,'' in *Proc. 3rd Conf. Elect. Comput. Eng. Technol.*, Tehran, Iran, 2017, pp. 1–6.

[6] Z. Raud and V. Vodovozov, ''Advancements and restrictions of E-assessment in view of remote learning in engineering,'' in *Proc. IEEE 60th Int. Sci. Conf. Power Electr. Eng. Riga Tech. Univ. (RTUCON)*, Riga, Latvia, Oct. 2019, pp. 1–6.

[7] T. Gorichanaz, ''Accused: How students respond to allegations of using ChatGPT on assessments,'' *Learn., Res. Pract.*, vol. 9, no. 2, pp. 183–196, Jul. 2023.

[8] N. C. Rowe, ''Cheating in online student assessment: Beyond plagiarism,'' *Online J. Distance Learn. Admin.*, vol. 7, pp. 1–8, Jan. 2004.

[9] J. Moten Jr., A. Fitterer, E. Brazier, J. Leonard, and A. Brown, ''Examining online college cyber cheating methods and prevention measures,'' *Electron. J. E-Learn.*, vol. 11, no. 2, pp. 139–146, 2013.

[10] A. Ullah, H. Xiao, and T. Barker, ''A classification of threats to remote online examinations,'' in *Proc. IEEE 7th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Vancouver, BC, Canada, Oct. 2016, pp. 1–7.

[11] C. Y. Chuang, S. D. Craig, and J. Femiani, ''Detecting probable cheating during online assessments based on time delay and head pose,'' *Higher Educ. Res. Develop.*, vol. 36, no. 6, pp. 1123–1137, Sep. 2017.

[12] S. Cerimagic and M. R. Hasan, ''Online exam vigilantes at Australian universities: Student academic fraudulence and the role of universities to counteract,'' *Universal J. Educ. Res.*, vol. 7, no. 4, pp. 929–936, Apr. 2019.

[13] T. M. Clark, C. S. Callam, N. M. Paul, M. W. Stoltzfus, and D. Turner, ''Testing in the time of COVID-19: A sudden transition to unproctored online exams,'' *J. Chem. Educ.*, vol. 97, no. 9, pp. 3413–3417, Sep. 2020.

[14] L. Elsalem, N. Al-Azzam, A. A. Jum'ah, and N. Obeidat, ''Remote E-exams during COVID-19 pandemic: A cross-sectional study of students' preferences and academic dishonesty in faculties of medical sciences,'' *Ann. Med. Surg.*, vol. 62, pp. 326–333, Feb. 2021.

[15] N. A. Shukor, Z. Tasir, and H. Van der Meijden, ''An examination of online learning effectiveness using data mining,'' *Proc.-Social Behav. Sci.*, vol. 172, pp. 555–562, Jan. 2015.

[16] N. Yan and O. T.-S. Au, ''Online learning behavior analysis based on machine learning,'' *Asian Assoc. Open Universities J.*, vol. 14, no. 2, pp. 97–106, Dec. 2019.

[17] I. D. Erguvan, ''The rise of contract cheating during the COVID-19 pandemic: A qualitative study through the eyes of academics in Kuwait,'' *Lang. Test. Asia*, vol. 11, no. 1, pp. 1–21, Dec. 2021.

[18] M. A. Sarrayrih and M. Ilyas, ''Challenges of online exam, performances and problems for online university exam,'' *Int. J. Comput. Sci. Issues*, vol. 10, no. 1, pp. 439–443, 2013.

[19] C. S. González-González, A. Infante-Moro, and J. C. Infante-Moro, ''Implementation of E-proctoring in online teaching: A study about motivational factors,'' *Sustainability*, vol. 12, no. 8, p. 3488, Apr. 2020.

[20] O. L. Holden, M. E. Norris, and V. A. Kuhlmeier, ''Academic integrity in online assessment: A research review,'' *Frontiers Educ.*, vol. 6, pp. 533–536, Jul. 2021.

[21] G. Kaisara and K. J. Bwalya, ''Strategies for enhancing assessment information integrity in mobile learning,'' *Informatics*, vol. 10, no. 1, p. 29, Mar. 2023.

[22] A. Wahid, Y. Sengoku, and M. Mambo, ''Toward constructing a secure online examination system,'' in *Proc. 9th Int. Conf. Ubiquitous Inf. Manage. Commun.*, Bali, Indonesia, Jan. 2015, pp. 1–8.

[23] A. Ullah, H. Xiao, and T. Barker, ''A dynamic profile questions approach to mitigate impersonation in online examinations,'' *J. Grid Comput.*, vol. 17, no. 2, pp. 209–223, Jun. 2019.

[24] S. S. Chua, J. B. Bondad, Z. R. Lumapas, and J. D. L. Garcia, ''Online examination system with cheating prevention using question bank randomization and tab locking,'' in *Proc. 4th Int. Conf. Inf. Technol. (InCIT)*, Bangkok, Thailand, Oct. 2019, pp. 126–131.

[25] M. S. Diarsini, L. P. Artini, N. N. Padmadewi, N. M. Ratminingsih, I. G. A. L. P. Utami, and N. P. E. Marsakawati, "Challenges and opportunities of online assessment implementation during COVID-19 pandemic in Indonesia based on recent studies," *Eur. J. Educ. Pedagogy*, vol. 3, no. 6, pp. 82–88, Nov. 2022.
[26] C. Rapanta, L. Botturi, P. Goodyear, L. Guàrdia, and M. Koole, "Online university teaching during and after the COVID-19 crisis: Refocusing teacher presence and learning activity," *Postdigital Sci. Educ.*, vol. 2, no. 3, pp. 923–945, Oct. 2020.
[27] G. Herrera, M. Nuñez-del-Prado, J. G. L. Lazo, and H. Alatrista, "Through an agnostic programming languages methodology for plagiarism detection in engineering coding courses," in *Proc. IEEE World Conf. Eng. Educ. (EDUNINE)*, Lima, Peru, Mar. 2019, pp. 1–6.
[28] N. M. Kingston and A. K. Clark, *Test Fraud: Statistical Detection and Methodology*. Evanston, IL, USA: Routledge, 2014.
[29] J. Ranger, N. Schmidt, and A. Wolgast, "The detection of cheating on E-exams in higher education—The performance of several old and some new indicators," *Frontiers Psychol.*, vol. 11, pp. 1–16, Oct. 2020.
[30] Y. Zhang, "Academic cheating as planned behavior: The effects of perceived behavioral control and individualism-collectivism orientations," *Higher Educ.*, vol. 87, no. 3, pp. 567–590, Mar. 2024.
[31] S. Prathish, A. N. S., and K. Bijlani, "An intelligent system for online exam monitoring," in *Proc. Int. Conf. Inf. Sci. (ICIS)*, Kochi, India, Aug. 2016, pp. 138–143.
[32] M. Wlodarczyk, D. Kacperski, P. Krotewicz, and K. Grabowski, "Evaluation of head pose estimation methods for a non-cooperative biometric system," in *Proc. 23rd Int. Conf. Mixed Design Integr. Circuits Syst.*, Jun. 2016, pp. 394–398.
[33] S. Hu, X. Jia, and Y. Fu, "Research on abnormal behavior detection of online examination based on image information," in *Proc. 10th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, Hangzhou, China, Aug. 2018, pp. 88–91.
[34] I. N. Yulita, F. A. Hariz, I. Suryana, and A. S. Prabuwono, "Educational innovation faced with COVID-19: Deep learning for online exam cheating detection," *Educ. Sci.*, vol. 13, no. 2, p. 194, Feb. 2023.
[35] C. F. Roger, "Faculty perceptions about E-cheating during online testing," *J. Comput. Sci. Colleges*, vol. 22, no. 2, pp. 206–212, 2006.
[36] G. R. Cluskey Jr., C. R. Ehlen, and M. H. Raiborn, "Thwarting online exam cheating without proctor supervision," *J. Academic Bus. Ethics*, vol. 4, no. 1, pp. 1–8, 2011.
[37] S. Patil, G. Norcie, A. Kapadia, and A. Lee, "'Check out where I am!': Location-sharing motivations, preferences, and practices," in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst.*, Austin, TX, USA, May 2012, pp. 1997–2002.
[38] F. Du, Y. Zhang, X. Bao, and B. Liu, "FENet: Roles classification of IP addresses using connection patterns," in *Proc. Int. Conf. Inf. Comput. Technol. (ICICT)*, Kahului, HI, USA, 2019, pp. 158–164.
[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[40] G. Huang, Z. Liu, G. Pleiss, L. van der Maaten, and K. Q. Weinberger, "Convolutional networks with dense connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8704–8716, Dec. 2022.
[41] N. Kriegeskorte and T. Golan, "Neural network models and deep learning," *Current Biol.*, vol. 29, no. 7, pp. R231–R236, Apr. 2019.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–15.

LESLIE CHING OW TIONG received the B.Sc. degree (Hons.) in computer science from Sunway University, Malaysia, in 2012, the M.Sc. (by Research) degree in computer science from Lancaster University, U.K., in 2014, and the Ph.D. degree in information and telecommunication technology (engineering) from Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea, in 2019. He is currently a Staff Engineer with Samsung Electronics, Republic of Korea. His research interests include deep learning, biometrics application, face recognition, periocular recognition in the wild, multi-view 3D reconstruction, and autonomous systems.

YUNLI LEE (Senior Member, IEEE) received the B.I.T. degree (Hons.) in software engineering from Multimedia University, Malaysia, in 2002, the master's degree in software from Dongseo University, South Korea, in 2004, and the Ph.D. degree in engineering (digital media) from Soongsil University, South Korea, in 2009. She is currently an Associate Professor with the Department of Computing and Information Systems, School of Engineering and Technology, Sunway University. She is also a Researcher with the Research Centre for Human-Machine Collaboration (HUMAC). She is also a Professional Technologist of MBOT and the Malaysia Director of the International Association for Convergence Science and Technology (IACST). Her current research interests include ultrasound imaging, the time series of FOREX data, technology modules for seniors, and augmented reality technology.

KAI LI LIM received the B.Eng. degree (Hons.) in electronic and computer engineering from the University of Nottingham, Nottingham, U.K., in 2012, the M.Sc. degree in computer science from Lancaster University, Lancaster, U.K., in 2014, and the Ph.D. degree in electrical and electronic engineering from The University of Western Australia, Perth, Australia, in 2020. He is currently the inaugural St Baker Fellow of E-Mobility with The University of Queensland, Brisbane, Australia. His research interests include visual navigation, navigational algorithms, and electric vehicles.

HEEJEONG JASMINE LEE received the B.Sc. degree in computer science from POSTECH, South Korea, the M.Sc. degree in computer science from the University of Edinburgh, U.K., the M.Phil. degree in technology policy from Cambridge University, U.K., and the Ph.D. degree in information technology from Monash University, Australia. She was with Korea Telecom Research and Development for eight years; and has filed 11 patents on telecommunications applications. Currently, she is a Research Professor with the College of Information and Communication Engineering, Sungkyunkwan University, South Korea. She moved from industry to academia was motivated by a love of learning. Her research interests include AI in dentistry and AI-semiconductor.