People's Democratic Republic of Algeria
الجمهورية الديمقراطية الشعبية الجزائرية
Ministry of Higher Education and Scientific Research
وزارة التعليم العالي والبحث العلمي
Chadli Bendjedid University - El Taref
جامعة الشاذلي بن جديد - الطارف
Faculty of Science and Technology
كلية العلوم والتكنولوجيا

THESIS

For the award of the Doctoral Degree (Third Cycle)

An Approach for Implementing E-learning Systems

Field: Computer Science

Specialization: Information Systems and E-Learning

Prepared By: Boutabia Ines

Defended on: 29/06/2026

Before the jury composed of:

Supervisor	Dr. Abdelmadjid BENMACHICHE	University of El-Taref
Co-Supervisor	Dr. Ali Abdelatif Betouil,	University of El-Taref
President	Dr. Ala Djeddai	University of El-Taref
Examiner	Dr. Faiz Maazouzi	University of Souk Ahras
Examiner	Dr. Amina Makhlouf	University of El-Taref

Academic Year: 2025/2026

Abstract

The advent of eLearning has witnessed exponential growth over the years, leading to the creation of new ways to make educational experiences more accessible, adaptable, and scalable. Nevertheless, a large number of the current systems continue to face the same problems of adaptability, engagement, and personalization. The idea of the Open Classroom is one of the solutions that addresses these issues by focusing on learner-centered and interactive environments. However, it still needs intelligent, data-driven models for its complete unfolding of potential. This thesis offers a solution to these demands through the proposal of a set of artificial intelligence frameworks that are mutually supportive of one another for the purpose of improving prediction, engagement, interaction, and information retrieval in the Open Classroom.

For this purpose, we present the Hybrid Reasoning Transformer with Gated Recurrent Unit (HRT-GRU), a model that predicts student progress more efficiently by combining contextual reasoning and sequential learning paradigms. Furthermore, a CNN-ViT hybrid system is proposed, which enhances student engagement detection by combining the local spatial data with the global attention mechanisms. Also, the T5-EPO-IoT virtual assistant is designed to make the learning process more flexible and interactive with the help of advanced natural language comprehension and optimization techniques. Finally, the LSTM-PSO intelligent search engine is launched, which makes the retrieval of educational materials more aware of the context and thus ensures efficient access to relevant content.

Broadly, these individual contributions represent a unified design that intends to promote the Open Classroom as an educational ecosystem that is more insightful, flexible, and personalized. Bridging major areas such as prediction, engagement, interaction, and information retrieval, this study overtly demonstrates the disruptive nature of artificial intelligence in the field of eLearning, thereby making the educational experience more efficient and diverse.

ملخص

شهد ظهور التعلم الإلكتروني نموًا هائلاً على مر السنين، مما أدى إلى ابتكار طرق جديدة لجعل التجارب التعليمية أكثر سهولةً ومرونة وقابليةً للتطوير. ومع ذلك، لا يزال عدد كبير من الأنظمة الحالية يواجه نفس مشكلات التكيف والمشاركة والتخصيص. تُعد فكرة الفصل الدراسي المفتوح أحد الحلول التي من شأنها معالجة هذه المشكلات من خلال التركيز على بيئات تفاعلية تركز على المتعلم. ومع ذلك، لا يزال هذا النظام بحاجة إلى نماذج ذكية قائمة على البيانات لإطلاق إمكاناته بالكامل. تقدم هذه الأطروحة حلاً لهذه المتطلبات من خلال اقتراح مجموعة من أطر عمل الذكاء الاصطناعي التي يدعم بعضها بعضًا بهدف تحسين التنبؤ والمشاركة والتفاعل واسترجاع المعلومات في الفصل الدراسي المفتوح.

ولهذا الغرض، نقدم نموذج محول التفكير الهجين مع الوحدة المتكررة المغلقة (HRT-GRU) ، وهو نموذج يتنبأ بتقدم الطلاب بكفاءة أكبر من خلال الجمع بين التفكير السياقي ونماذج التعلم التسلسلي. علاوةً على ذلك، يُقترح نظام هجين يجمع بين CNN وViT، يُعزز رصد تفاعل الطلاب من خلال دمج البيانات المكانية المحلية مع آليات الانتباه العالمية. كما صُمم المساعد الافتراضي T5-EPO-IoT لجعل عملية التعلم أكثر مرونةً وتفاعليةً بفضل تقنيات فهم اللغة الطبيعية المتقدمة وتحسينها. وأخيرًا، تم إطلاق محرك البحث الذكيLSTM-PSO ، مما يجعل استرجاع المواد التعليمية أكثر وعيًا بالسياق، وبالتالي يضمن وصولًا فعالًا إلى المحتوى ذي الصلة.

بشكل عام، تُمثل هذه المساهمات الفردية تصميمًا موحدًا يهدف إلى تعزيز الفصل الدراسي المفتوح كنظام بيئي تعليمي أكثر ثراءً ومرونة وشخصية. من خلال ربط مجالات رئيسية مثل التنبؤ والمشاركة والتفاعل واسترجاع المعلومات، تُظهر هذه الدراسة بوضوح الطبيعة الثورية للذكاء الاصطناعي في مجال التعلم الإلكتروني، مما يجعل التجربة التعليمية أكثر كفاءةً وتنوعًا.

Résumé

L'avènement de l'apprentissage en ligne a connu une croissance exponentielle au fil des ans, ouvrant la voie à de nouvelles façons de rendre les expériences éducatives plus accessibles, adaptables et évolutives. Néanmoins, un grand nombre de systèmes actuels restent confrontés aux mêmes problèmes d'adaptabilité, d'engagement et de personnalisation. L'idée de la classe ouverte est l'une des solutions pour répondre à ces enjeux en privilégiant des environnements interactifs et centrés sur l'apprenant. Cependant, elle nécessite encore des modèles intelligents, basés sur les données, pour déployer pleinement son potentiel. Cette thèse propose une solution à ces exigences en proposant un ensemble de cadres d'intelligence artificielle complémentaires afin d'améliorer la prédiction, l'engagement, l'interaction et la recherche d'informations dans la classe ouverte.

À cette fin, nous présentons le Transformateur de raisonnement hybride avec unité récurrente fermée (HRT-GRU), un modèle qui prédit plus efficacement les progrès des élèves en combinant raisonnement contextuel et paradigmes d'apprentissage séquentiel. De plus, un système hybride CNN-ViT est proposé, qui améliore la détection de l'engagement des étudiants en combinant les données spatiales locales avec les mécanismes d'attention globaux. De plus, l'assistant virtuel T5-EPO-IoT est conçu pour rendre le processus d'apprentissage plus flexible et interactif grâce à des techniques avancées de compréhension et d'optimisation du langage naturel. Enfin, le moteur de recherche intelligent LSTM-PSO est lancé, rendant la recherche de ressources pédagogiques plus contextuelle et garantissant ainsi un accès efficace au contenu pertinent.

Ces contributions individuelles constituent un projet unifié visant à promouvoir l'Open Classroom comme un écosystème éducatif plus perspicace, flexible et personnalisé. En reliant des domaines majeurs tels que la prédiction, l'engagement, l'interaction et la recherche d'information, cette étude démontre clairement le caractère disruptif de l'intelligence artificielle dans le domaine de l'apprentissage en ligne, rendant ainsi l'expérience éducative plus efficace et diversifiée.

To the loving memory of my Beloved Father,

Sadreddine Boutabia

I wish you were here to witness this moment, but I carry you with me in every step I take and make sure you are proud of what you left behind

ALWAYS.

Acknowledgment

I would like to express my deepest and greatest gratitude to my supervisor, mentor, and friend Professor **Abdelmadjid Benmachiche**, whose mentorship has been instrumental in helping me reach this milestone. His continuous support, expert guidance, and patience have shaped my academic growth and provided the direction needed to successfully complete this work.

I extend my heartfelt thanks to my co-supervisor, Dr. **Ali Abdelatif Betouil**, for his assistance and constructive feedback. His expertise and dedication have significantly contributed to the progress and completion of this thesis.

I would also like to extend my sincere appreciation to the president and the members of the academic committee for accepting the invitation today, as well as to all the teachers and staff who have contributed to my education and honored me with their presence and support. Their insights, evaluation, and dedication to academic excellence are deeply appreciated.

I am also profoundly thankful to my family, my mother, and my sister **Abir** and my siblings for their unconditional love, understanding, and constant encouragement. Their belief in me has been my greatest strength.

My sincere thanks go to my friends, and 한국어 수업 친구들 **Imen, Afef, and Abdelkader**, who have stood by me and endured my stress, nagging, and mood swings with kindness and humor. Your support and companionship made this journey much more bearable.

To everyone who contributed in one way or another to this achievement, I am truly grateful.

Sincerely

INES BOUTABIA

Table of Content

Table of Figures

List of Tables

GENERAL INTRODUCTION

Over the last twenty years, digital tools have transformed the way knowledge is acquired and learning via eLearning, which applies internet connectivity, digital devices, and online platforms to facilitate learning. At the beginning, eLearning revolved around computer-assisted instruction with somewhat basic interactive features; however, it has grown into a complex system including MOOCs [1], virtual classrooms, and AI-based customized learning.

E-learning is compatible with both synchronous and asynchronous formats; it opens the door to a live session or a self-paced course. Besides, the use of gamification, virtual reality, and augmented reality has attracted more students and provided visitors with fascinating and immersive ecosystems to learn. It is a worldwide phenomenon that, among other things, allows educational experiences to be scaled up or down, to be more flexible and interactive, and enables the entire world population to have access to high-quality education. eLearning has eliminated the educational barrier of "location" or "budget" and is therefore the ultimate way to ensure learning continuity and development of skills in the era of digitalization.

Multimedia-rich content, gamified interactive assessments, and robust communication tools are some of the key features of modern eLearning systems. Besides these, smart features like recommendation engines and adaptive learning modules are also being increasingly used to provide tailor-made solutions for each learner. These systems are being used in different educational settings like schools, universities, corporate training programs, and informal self-paced learning environments. They are also packed with a range of benefits. They can be accessed from anywhere, regardless of the geographical location; learners can learn at their own pace if they wish. Furthermore, they are an affordable educational solution and feature continuous assessment and feedback.

On the other hand, these advantages notwithstanding, the platforms of eLearning usually have a lot of drawbacks. A majority of them still employ traditional content delivery methods, which are not conducive to providing the needed personalization to create individualized learning paths. Additionally, they are often somewhat marginal in their lack of use of modern artificial intelligence, IoT, and big data analytics. As a result, the learning journey may be less vivid, less attractive, and at times less efficient compared to what could be. Numerous contemporary eLearning problems include inadequate personalization, ineffective engagement tracking, slow resource retrieval processes, and minimal real-time adaptability. Therefore, although digital learning milieus have become more user-friendly than at any time in the past, their effectiveness in facilitating learning gains is still varying.

Yet, despite significant virtues, eLearning platforms are frequently fraught with severe defects. Many of them are still heavily dependent on static content delivery methods, which not only fail to meet the personalization needs of learners but also do not provide individualized learning paths. In addition, they are often nowhere near fully exploiting the potential of modern artificial intelligence, IoT, and advanced data analytics. Therefore, the process of learning can be less lively, less attractive, and at times less efficient. Current eLearning facilities often face challenges, such as a lack of personalization, insufficient monitoring of engagement, slow access to resources, and limited adaptability in real-time. As a result, although digital learning environments have become more accessible than ever before, their effectiveness in improving learning outcomes remains unstable.

Within this framework, the Open Classroom (OCR) Initiative is a new, innovative, and promising approach to education to develop a more vibrant and learner-centric educational environment. Through such an initiative, the classrooms enable students to be involved in different interactive, resourceful, and flexible spaces, such as a traditional physical classroom, an immersive virtual environment, or hybrid ones that combine the two, transcending the limitations of each. The primary goal of the plan is, however, to build an attractive environment that encourages interaction among the learners, triggers their creative ideas, and also helps them to learn actively. Nevertheless, the Open Classroom Initiative, which is able to reshape educational journeys, still encounters some difficulties.

Among the most critical issues are the need to efficiently customize educational materials that are suitable for each learner, the ability to smartly monitor and forecast student progress over time, and the necessity of maintaining a high level of engagement throughout the lessons. Moreover, it is also important that access to learning resources is meaningful in terms of the learning context and that it is in accordance with the student's unique learning objectives. Non-compliance with these issues would lead to OCR being at the same level or even worse than the existing eLearning methodologies, resulting in the loss of educational transformation potential and the future education vision.

This dissertation proposes a comprehensive set of AI- and IoT-powered solutions designed explicitly to supplement the OCR Initiative to overcome these constraints. The novel first contribution goes to the HRT-GRU, a hybrid architecture that melds a Hierarchical

Reasoning Transformer with a Gated Recurrent Unit. This model, by implementing complex data patterns, sets out to give accurate forecasts of student progress. The other contribution is a CNN-ViT model that amalgamates Convolutional Neural Networks with Vision Transformers for a straightforward and fast engagement detection process; thus, a dynamic evaluation of student participation can be made. A T5-EPO-IoT virtual assistant that merges the features of the T5 Transformer model with Emperor Penguin Optimization and IoT sensor technology for a continuous, interactive, and adaptive learning support customized for each individual is the third part. The LSTM-PSO smart search engine, which integrates Long Short-Term Memory networks with Particle Swarm Optimization to enable context-aware and personalized educational resource retrieval, thus, learners acquire the most relevant materials correlated with their unique objectives, is the last one. As a whole, these contributions aim to make the OCR an intelligent, adaptive, and responsive environment that is capable of providing a more enriching and effective educational experience.

This thesis comprises four detailed chapters. Chapter 1 contextualizes eLearning and the Open Classroom Initiative, outlining their evolutionary history, defining the basic elements, acknowledging numerous usages, and providing a balanced opinion of advantages and limitations. Additionally, the chapter also features the integration of AI and IoT technologies in education that results in better learning frameworks. Chapter 2 contributes to the present work by providing a detailed survey of the literature on intelligent search engines, virtual learning assistants, engagement detection devices, and prediction models for the student's performance based on the setting of the educational context. Chapter 3 describes the suggested methods in detail, the four primary approaches proposed as part of this research: predicting student progress, a system for detecting engagement, the virtual assistant (VA) for adaptive learning support, and the intelligent search engine for tailored resource retrieval. The experimental procedures, results, and discussions for each system, which focus on performance, unique advantages, and possible improvements of the systems, are also included in this chapter. Finally, the research work discussed in Chapter 4 draws together the main conclusions, providing an overview of the primary contributions made by this study, reflecting on their implications for the future of education, and indicating possible directions for future research and development in this area.

Chapter 1: Foundations and Evolution of E-Learning

1. Introduction

In the previous chapter, we provided the general introduction to this thesis, offering an overview of the major changes the educational landscape has experienced due to the digital era. This overview covered the technological, pedagogical, and societal shifts that have impacted the education sector. We discussed the transition in the teaching paradigm from traditional classroom settings to virtual or hybrid learning modes, emphasizing that these changes have introduced new ways that are transforming the interactions between educators, students, and educational materials.

This chapter, grounded on that, examines the drift and conceptual change of E-Learning in detail. We will trace how the development of E-Learning has changed from a basic distance education method to the present complex and interactive learning environments, where Artificial Intelligence (AI) and adaptive learning technologies play a significant role [2]. The presentation of educational theories that have been most influential in the formation of e-learning systems will be accompanied by the recognition of the main technological breakthroughs, such as cloud computing and m-learning, that have expedited their growth and accessibility

Furthermore, we will explore the emergence and coexistence of smart and open systems in the contemporary digital classroom, signifying the upheaval that such discoveries have prompted in the features and design of educational tools, especially the new system of education dealt with in this dissertation.

Next, we will thoroughly examine the OCR model, understanding and explaining its basic principles, the system's structural features, and its essential role in enabling an environment for E-Learning systems to function. By establishing a clear link between E-Learning development and the OCR idea, our goal is to create a strong theoretical and technological foundation. This base will support us in tracing the advantages and disadvantages of the existing approach, thus making us able to suggest the solutions that not only suit the challenges but also make use of the untapped potentials of educational technology.

2. E-Learning: Foundations and Evolution

2.1. The Concept of E-Learning

The term "E-Learning" is commonly linked to Elliott Masie [3], who first introduced it in 1999 at the TechLearn Conference [4]. He presented it as a groundbreaking method that uses the web to provide training and schooling beyond the closed walls of a traditional classroom. Later on, the phrase "E-Learning" was changed to a single comprehensive term that refers to various all-in-one learning experiences through different electronic technologies.

Fundamentally, E-Learning is a revolution in the field of education, where the use of digital tools and resources is a major and extremely successful aspect of the teaching and learning process. This development represents a departure from the static, traditional educational models, which treated all students alike, and is still very prominent in the past, but E-Learning has set a new standard that is more flexible and learner-centered. In this type of setup, learners acquire skills to manage their own pace of learning, choose paths freely, and, in some instances, personalize their materials, thereby making their interaction with the topic deeper. E-Learning is not limited to physical or temporal boundaries like the traditional classroom; therefore, it can be found everywhere and can be extended to a larger number of people [5].

The idea of E-Learning is built around three main areas of concern:

- ***Mediation by Technology:*** All learning activities are facilitated through various electronic means such as computers, smartphones, tablets, or cloud-based platforms, though not limited to these only. These cutting-edge technologies open up new routes for content delivery, customer engagement, and measuring the level of understanding.
- ***Digital Content and Interaction:*** E-Learning is an extension of conventional education, where lectures and textbooks are the primary mode of teaching and learning. A cornucopia of multimedia resources, including videos, podcasts, interactive simulations, and gamified elements, is the mainstay of E-Learning that not only deepens the subject matter but also makes it more accessible to different learner styles and preferences.
- ***Learner Autonomy and Accessibility***: E-Learning stands out particularly in its potential to enable learners. People can access educational materials at their

disposal whenever they want, and in a way that suits their flexible schedules and learning preferences. The upshot of this leveling of the education field is the tremendous growth in accessibility, which, in turn, ensures that a broad audience, endowed with quality learning experiences irrespective of their geographical or socio-economic backgrounds, is the ultimate beneficiary.

E-Learning, notably, is not solely confined to distance learning formats. Moreover, the use of technology in instruction, along with the blended learning approach, provides students with a better understanding of the subject. Furthermore, the AI and data analytics employed in adaptive learning methods to customize educational activities for each learner make the journey even more personal.

E-Learning was one of those times in educational discourse when the new term was not simply a new word to describe something familiar, but rather a new concept. It indicated a change of the paradigm in education, a recasting of learning as an ongoing, flexible, and non-closed class-based process. To use the phrasing of the article, E-Learning 'breaks boundaries' not only because it improves the availability of education but also because it opens up the opportunities of ongoing and lifelong learning for people of different ages.

2.2. Historical Progression of E-Learning

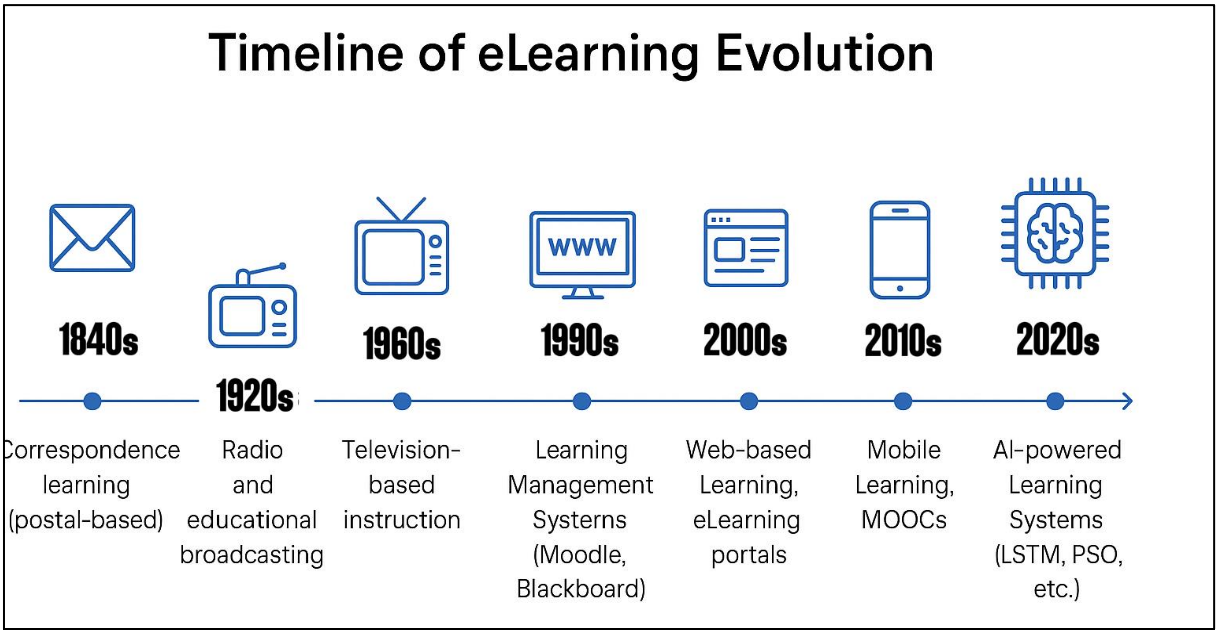

Figure 1-1: E-learning Evolution timeline

The transformation of E-Learning has always been dependent on evolutions in technology and the development of new teaching methods. Initially, e-learning was only possible with programmed instruction, whereas today it can still be seen that the educational sector has made its way through numerous revolutionary phases in which e-learning has adapted its response to the technological and pedagogical trends.

- ***1840s – Correspondence Learning***

In the 1840s, the introduction of correspondence learning was a major change in the education system. The new method was a distance learning system that used the post to communicate. Printed materials were sent to the students, which included textbooks, assignments, and instructional guides. They would do the assignments in their free time and then send their work back to the instructors by mail for grading. Even though it was a simple system, correspondence education efficiently distinguished the teacher and the learner in both time and space, thus setting an important precedent for other distance education modalities. With this method, people who lived in far-off places or had little access to the normal educational institutions were allowed to study on their own.

- ***1920s– Educational Broadcasting***

Between the 1920s and the late 1940s, the rise of educational broadcasting through radio and television led to a big change in education [6]. Practically any family could access schools and colleges through the media mentioned above nationwide. Radio schools provided lectures and language courses to those who wanted to learn something new, whereas educational TV attracted more people since it allowed the visualization of the subjects taught. The era represented a one-to-many communication model, where an individual teacher had many students but very little or no interaction; hence, it laid the foundation for later interactive educational formats.

- ***1950s – The Beginning of Programmed Instruction and Teaching***

The development of E-Learning can initially be linked to B.F. Skinner's experimental "teaching machines" were created in the mid-20th century [7]. The first machines were examples of programmed instruction, which enabled learners to interact with a logically arranged sequence of questions and thus get immediate feedback on their answers. At the same time, these devices were very basic and mechanical, yet they were the first step towards self-paced, technology-mediated learning, which was a change in the concept of the teacher's presence in the classroom.

- ***1960s–1970s – Computer-Assisted Instruction (CAI)***

During the time when mainframe computers were becoming more widespread, Computer-Assisted Instruction began to prosper in the educational and research departments. One of the remarkable achievements was the creation of PLATO (Programmed Logic for Automatic Teaching Operations) at the University of Illinois in 1960 [8]. The PLATO project introduced new elements, including asynchronous forums, message boards, and online testing functionalities; these features were already present in modern online learning platforms. However, the use of CAI as interactive learning was still largely limited to those institutions that had the financial resources to support the computing infrastructure necessary for such a costly implementation [9].

- ***1980s – Computer-Based Training (CBT)***

The creation of personal computers in the 1980s significantly altered the course of E-Learning [10]. One of the major changes was the rise of multimedia learning programs on CD-ROMs and floppy disks. This was especially important in business training environments. Throughout this period, the major feature of the programs was the possibility of performing the training at one's own pace, and the modules were designed for vocational training and skill development. They also allowed employees to become proficient in the areas they preferred by using the unique method that they learned. Despite this, CBT was mostly carried out in standalone offline environments.

- ***1990s – The Internet and the Birth of E-Learning***

One of the most significant technological advancements that changed the entire E-Learning landscape was the internet. Consequently, E-Learning was transformed into an online-based activity that could be accessible anywhere, irrespective of physical distance. In 1999, the phrase "E-Learning" was invented by Elliott Masie, implying a radically different way of teaching/exchange of knowledge. Besides facilitating more structured online course delivery, the advent of Learning Management Systems (LMSs) [11] like Blackboard and Moodle also allowed the scalable distance education era to be initiated. This decade reflected the use of various multimedia components, for instance, video lectures, interactive quizzes, and discussion forums, which not only allowed students to learn but also to practice and discuss their misconceptions.

- ***2000s – Web 2.0 and the Rise of MOOCs***

The transition to a more user-centered web, known as Web 2.0, had a massive change on the E-Learning sector [12]. The advent of social media made it possible for students not only to actively consume content but also to create, in collaboration with other students, shared projects through wikis, and get involved in discussions via blogs or forums. The period witnessed the emergence of Massive Open Online Courses (MOOCs) [13], among which are such platforms as Coursera [14], edX [15], and Udacity [16] are the most recognized. They not only abolished the traditional walls of academia but also made it possible for anyone from anywhere to take part in the courses of the world’s best universities. Even though these developments provided easy access to knowledge for everyone, they raised questions about the number of students completing courses and the quality of education [17].

- ***2010s – Mobile Learning, Cloud Computing, and Analytics***

In this decade, the widespread use of smartphones and tablets changed the way people learn, offering the opportunity to access education "anytime, anywhere". Cloud-based LMS solutions allowed for easy and efficient delivery of educational content, which made learning more adaptable to students' needs. Furthermore, the emergence of Learning Analytics empowered educators to use data for tracking the progress of students, their level of commitment, as well as their overall performance. Along with this, the period also saw the advent of the use of gamification elements, by employing characters and storylines of games to learning and microlearning techniques, which matched with learners' tastes for brief and easy-to-understand content [18].

- ***2020s – AI, IoT, and Smart Learning Environments***

The COVID-19 pandemic ultimately became a significant promoter of E-Learning adoption all over the world. As a result, E-Learning went from being a supplementary choice to a mainstream educational necessity. Present-day E-Learning systems gradually incorporate AI [19], as it enables adaptive learning, which adjusts educational experiences to the competency level of an individual. Moreover, Natural Language Processing (NLP) is the technology that enables virtual assistants, which help learners in real-time, whereas the

Internet of Things (IoT) is regarded as a factor that leads to the design of groundbreaking and interconnected classroom environments.

At this time, the use of AI-powered recommendation systems and intelligent tutoring platforms has become widespread [20], as well as blended learning frameworks that enable the use of online resources to be combined with face-to-face instruction. The emergence of the Open Classroom (OCR) Initiative is becoming a logical step forward; such a transformation would imply a shift in E-Learning towards more flexible, interactive, and intelligent educational ecosystems that put learner engagement and personalized learning paths first.

2.3. Components of E-Learning

The success of E-Learning heavily depends on a seamless combination of digital-technical infrastructure, instructional design, and tools that can engage the learner, which, in essence, are all interconnected elements of a digital learning ecosystem. These building blocks can be categorized into the following major areas: technological platforms, the ways of content creation and delivery, communication and collaboration tools, assessment systems, and advanced personalization features. Moreover, these features should play a significant role in guaranteeing that learning through e-education becomes a viable option for all and further that it results in an effective, efficient, and engaging learning experience.

2.3.1. Learning Management Systems (LMS)

Learning Management Systems (LMS) are the core framework of most E-Learning environments [11]. An LMS is a software suite that consolidates the distribution, administration, and monitoring of learning activities. Some of the most popular platforms, like Moodle [21], Blackboard [22], and Canvas [23], have gained wide application because of their strong features; these allow instructors to upload educational materials with minimum effort, manage enrollment of students, check both individual and group progress through different assessment activities, and evaluate the performance by using various metrics.

Moreover, these platforms enable cataloging classes with features like structured modules, guided learning paths, and user-friendly dashboards, which lead the learners to the logical study of the material. In addition, the emergence of cloud-based LMS solutions has profoundly impacted the sector, thereby offering institutions the choice of scalable options, which enable real-time updates and a variety of integrations with third-party tools. This change has made the deployment in academic and corporate learning environments more extensive, thus providing more flexibility and accessibility.

2.3.2. Content Delivery and Authoring Tools

Learning success is largely dependent on content quality, interactivity, and engagement levels. The use of authoring tools such as Articulate 360, Adobe Captivate, and H5P allows educators to create multimedia-rich learning modules that integrate text, images, audio, video, simulations, and gamified interactions without inconvenience. This not only makes the education more engaging but also easier to understand. The implementation of standards such as SCORM (Sharable Content Object Reference Model) [24] and xAPI (Experience API) [25] plays a vital role in assuring the interoperability of content among various platforms; thus, they can be used in different educational contexts with fewer restrictions.

Consequently, with the growing demand for mobile-first learning solutions, the content delivery method has shifted to a fully responsive design. Now learners are able to access the same materials on their different gadgets, ranging from smartphones, tablets, and laptops; therefore, no learning is jeopardized by the change of device.

2.3.3. Communication and Collaboration Tools

One of the biggest problems in online learning is still having the social aspect that is present in traditional classroom settings. Various communication tools such as discussion forums, instant messaging systems, and video conferencing platforms have emerged as essential components of E-Learning to bridge this gap. Top tools such as Zoom, Microsoft Teams, and Google Meet are among the leading platforms that enable synchronous learning, allowing instructors and learners to engage in live interactions.

Additionally, the implementation of asynchronous discussion channels, like forums and chatrooms, has facilitated communication and collaboration between those engaged in live sessions and those not. The use of collaborative tools, such as shared whiteboards, group project management applications, and peer-to-peer interaction networks, has become increasingly widespread, enabling learners to maintain continuous interaction even in virtual environments.

2.3.4. Assessment and Feedback Systems

Assessment remains one of the most vital parts of the educational process, and E-Learning platforms have different intricate systems to evaluate the performance of students. Online tests, assignments, and exams can be automatically graded for students, thus providing them with instant feedback, an essential part of their learning progression. The more sophisticated E-Learning platforms have begun to employ adaptive testing algorithms that, in real-time, select the difficulty level of the next question based on the previous answers of the learners, thereby offering a more personalized assessment experience.

Moreover, instructors benefit from performance analytics tools that monitor changes in performance over time, identify learners who may be struggling, and recommend suitable intervention at the right time and place. Furthermore, the feedback loop in E-Learning is not limited to grading only. Peer review tools and the detailed comments of instructors constitute the main element in the establishment of an ever-repeating cycle of improvement culture.

2.3.5. Personalization and Learning Analytics

Today's E-Learning paradigm is gradually shifting to the point that it is largely based on learner-centered approaches, which are substantially supported by personalization and learning analytics. By engaging an individual learner, the learning path can be adjusted to fit the distinct needs, interests, and performance levels of that learner. Among the intelligent systems, the recommendation system is the primary factor [26], as it always suggests the most relevant resources to the learner based on previous engagement and the learner's progress.

Learning analytics scrupulously gathers and depicts the data on the behavior of the learner, for example, time devoted to tasks, completion rates, and levels of engagement, thus providing both instructors and learners with actionable insights. This approach, driven by data, not only enhances the overall learning experience but also enables institutions to make decisions regarding curriculum design and delivery that are more informed and based on evidence.

2.3.6. Supporting Infrastructure

Stable infrastructure is the foundation of efficient virtual learning, in which it is not only software but also the necessary hardware, like a high-speed connection to the internet, safe cloud storage, well-protected security measures [27], and accessibility to various types of devices. Part of a successful educational environment is the provision of a high-speed internet connection and reliable server capabilities to allow smooth access to the learning materials.

Alongside this, the configuration of security protocols also protects the confidential information of users, guaranteeing both privacy and adherence to laws and regulations. The accessibility standards, such as the Web Content Accessibility Guidelines (WCAG) [28], also help in making E-Learning platforms accessible for learners with disabilities through the provision of features such as screen readers, subtitles, and customizable interfaces, thus enabling all learners to be part of the learning process.

2.4. Advantages of E-Learning

E-Learning has revolutionized education to the extent that it is unarguably one of the most impactful evolutionary changes, with different benefits attached to it that cater to the needs of both learners and educators. Not surprisingly, the benefits of this technology are not limited to either place or time, and thus, it changes the very way knowledge is acquired, transferred, and accumulated. E-Learning main points are described below:

- ***Accessibility and Flexibility***

One of the most notable benefits of E-Learning is the ability to access educational material anytime and from any place. The adaptability of this method frees the learners from the limitations of the traditional classroom setting and strict schedules. Students from remote areas or less privileged conditions can participate in courses from top universities that they have little or no access to. Consequently, employees can continue their learning without interruption; thus, maintaining their careers and sharpening their skills simultaneously. Such incredible accessibility underscores the need for an inclusive and lifelong learning approach, which is the current educational trend that adapts to different learning needs and various types of people.

- ***Cost-Effectiveness***

Traditional educational methods require substantial expense on infrastructure, printing, and logistics. However, E-Learning is a more affordable option. The cost of creating digital content can be spread among a large number of users, which significantly lowers the cost per learner. Moreover, E-Learning saves money for schools and students in the form of expenses on trips, lodging, and the purchase of textbooks. The lower prices not only facilitate access to education but also attract a wider range of participants, especially those who might not have enrolled due to a lack of funds.

- ***Personalization of Learning***

E-Learning is a completely different experience from the traditional one-size-fits-all approach, as it allows personalized learning experiences. The students can design their courses with the help of adaptive learning technologies, smart tutoring systems, and thorough learning analytics. They can also progress at their own pace, redo difficult areas again, or move faster through the material they already know, which results in a more involved and enthusiastic learning environment. This one-on-one interaction with the content not only engages the learner more but also facilitates the retention of knowledge since learners become stronger and feel more responsible for their learning journeys.

- ***Interactivity and Engagement***

Multimedia elements such as videos, animations, interactive simulations, and gamified content have changed E-Learning into an engaging mode of education. These vivid methods engage several senses, thus making comprehension easier and memory retention longer than when using traditional methods of passive learning, for instance, lectures. Moreover, E-Learning platforms, through the use of game features, can support learners who gain strength and will be more active if they acquire badges, rewards, and leaderboards, thus transforming the learning experience into a continuous and enjoyable way.

- ***Scalability and Reach***

The capability to expand E-Learning platforms is another important advantage that enables schools to meet the needs of a large number of students simultaneously without being restricted by the size of physical classrooms. This feature has been the main driver of the spread of MOOCs, which have changed the entire world's reach of high-quality education. Consequently, students with varied backgrounds can take advantage of expert-led courses, thus not only cross the distance and economic divide but also enable the educational sector to be more democratic.

- ***Continuous Learning and Lifelong Development***

The modern world is transforming at an Accelerated rate, and so is the knowledge economy. The need for professionals to constantly upgrade their skills to avoid obsolescence is therefore imperative. In this context, the role of E-Learning in the scene of continuous professional development cannot be overemphasized. E-Learning provides the opportunity for people to acquire certifications, refresh their technical skills, and become acquainted with new technologies without interrupting their career progression. Professionals afraid they might become irrelevant in their areas of practice now have a global reason to be optimistic, as E-Learning, by facilitating continuous upskilling and reskilling, is in perfect alignment with the global focus on lifelong learning.

- ***Rich Data and Analytics for Improvement***

Digital E-Learning platforms generate extensive amounts of data on learner tendencies, improvement, and achievement, thus providing educators with valuable insights. One of the benefits of learning analytics is the ability to detect students in trouble, therefore, provide prompt intervention. Furthermore, the adjustment of the content by the educators is possible with the help of teaching strategies. Schools have the opportunity to use this approach based on data to create the best methods of teaching and make the management of the institution more efficient, which will consequently foster better educational outcomes and institutional efficiency.

2.5. Disadvantages and Limitations of E-Learning

E-Learning has a lot of benefits. However, it also has various obstacles and restrictions that, in particular, may diminish the effectiveness of E-Learning in certain situations. It is essential to recognize these deficiencies to develop a better educational system.

- ***Lack of Human Interaction and Social Presence***

Reduced communications, or rather, the almost non-existence of face-to-face talks, is a major complaint from users of E-Learning regarding the latter as one of its drawbacks. In the absence of these social cues, such as body language and intonation, the students may feel lonely. This feeling of alienation can lead to a decrease in motivation and intention to participate, therefore, making it difficult for learners to establish new friendships and receive prompt replies required from teachers, as in the case of traditional teaching courses.

- ***Digital Divide and Accessibility Issues***

The availability of internet services that can be relied upon, digital devices, and the necessary digital literacy skills are distributed unequally in different areas and between various socio-economic groups. This difference in distribution creates an educational inequality in which certain populations are de facto excluded from valuable learning opportunities because of technological limitations. Therefore, students who do not have these resources may find it difficult to keep up with their classmates, which in turn will increase the existing educational disparities.

- ***Self-Discipline and Motivation Challenges***

E-Learning is inherently dependent on the personal responsibility of the learner, efficient time management, and the possession of the inner drive. Many students will suffer from putting off tasks and may, at last, cease their online studies due to a lack of a specific routine, and no one is there to hold them accountable immediately. Those who are Accustomed to operating within a certain framework in a regular classroom might struggle with concentration and enthusiasm if they have to study on their own without any direction.

- ***Quality and Credibility of Content***

The widespread use of online learning platforms has significantly diversified the quality of the content. Without any standardized oversight and accreditation, the academic rigor in these courses may even be lacking, thereby, the educational experience may be compromised. Furthermore, the continuous change of knowledge in different fields means that online material should be updated quite frequently to stay relevant; however, a majority of the courses are not able to fulfill this essential requirement.

- ***Technical Difficulties and Learning Curve***

Learners often encounter a variety of technical issues, including system errors, software compatibility challenges, and difficulties in using digital platforms. These hurdles can become Critical obstructions to acquiring knowledge and can be very annoying for the less tech-savvy individuals. Moreover, instructors might also require extensive training to be able to efficiently create, present, and administer digital content as well as attract their students through online channels.

- ***Assessment Limitations***

Online assessments have several limitations that include susceptibility to dishonesty and academic misconduct, which may lead to the inaccurate evaluation of learner performance. Moreover, practical skills, which are pivotal in sectors that require hands-on training, are difficult to be completely transferred into digital format; thus, learners will be only partially trained in such skills.

- ***Reduced Development of Soft Skills***

Online learning purely through the internet usually cannot develop key soft skills like teamwork, communication, and interpersonal interaction. Generally, these abilities are acquired in the traditional school setting through interaction among peers and involvement in group activities; however, these may be limited or even missing in an online learning format. Consequently, students might complete their studies lacking some crucial competencies.

3. The Open Classroom Initiative: Concept and Emergence

The OCR Initiative signifies a fundamental redefinition of the educational system that deepens and enhances the foundational features of E-Learning by far. Whereas the latter mainly concentrates on the mere digitalization and distribution of educational content via various tech channels, the values of the OCR are accessibility, inclusiveness, and interaction as its key features. The movement recognizes these three as the main features of the new educational paradigm after having crossed borders, put the learner first, and adapted the context. Consequently, it intends to break the barrier between schools and community learning, thus building the educational ecosystem that is both holistic and dynamic.

The development of the "open classroom" idea is recognizable from the late 60's and 70's, a time when the waves of the progressive educational movements in the USA and the UK had a significant impact on the concept. Initially, the concept mainly referred to new designs of the physical classroom that promoted a rethinking of educational space, i.e., the removal of the traditional walls allowed for more flexible environments in which students could engage, collaborate, and learn at their own paces. With the further development of these concepts, the idea of the OCR changed from a simple physical arrangement to a more complete pedagogical and technological framework. In this wider sense, openness did not limit itself to space only, but it also meant unbarred access to knowledge, the availability of Open Educational Resources (OER) [29], more inclusive participation mechanisms, and a range of open-ended learning experiences.

The OCR Initiative in the 21st century is a perfect instance of the digital transformation in the educational field. Essentially, this program embodies the scalability found in e-learning formats, the collaborative aspect that social learning systems have enabled, and the adaptability given by technological developments such as AI and IoT. Such an initiative not only offers a technologically advanced educational experience but also ensures pedagogical inclusiveness and the learners' engagement. By concentrating on the fundamentals of openness, the OCR Initiative is certain that learning opportunities are beyond the limits of traditional institutions, areas that are far apart from each other, and socio-economically disadvantaged communities.

3.1. Historical Progression of the OCR Initiative

Through a series of evolutions, the OCR Initiative has progressed from its original concept, a shift from the traditional classroom environment to modern digital and AI-assisted learning ecosystems. The transformation, of course, over time from the confined setting of physical classrooms to the present-day online, AI-augmented learning environment dates back many years and can be traced with the help of a few important events in the last century.

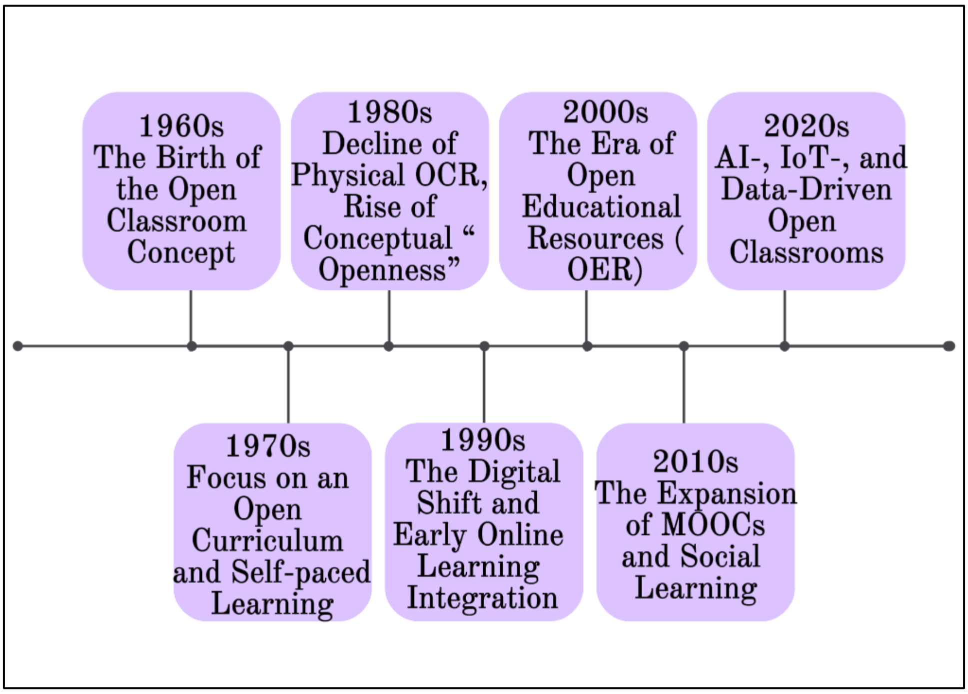

Figure 1-2: Historical Progression of the OCR Initiative

- ***1960s–1970s: The Birth of the Open Classroom Concept***

The term "open classroom" was a major feature of the US and UK scene during this time, greatly energized by the ideas of the progressive education movement, especially the views of

John Dewey on learning by experience and social reform through education. One of the first examples of the OCR idea was an open-plan school, where there were no walls or separations between classrooms. This change in school building was meant to raise the qualities of collaboration, autonomy, and flexibility among learners; thus, in practice, it was a break from the conventional teacher-centered pedagogy that was strict and inflexible. Normally, in such situations, students were given the power to explore their own methods of learning, choose different activities, and even learn from peers. Teachers shifted to the role of mentors, educating through guidance instead of directly instructing. Despite that, the OCR setup still had some issues in the real world, like increased noise levels, a shortage of organized learning pathways that could confuse students, and the problem of managing classrooms. These problems, in the end, limited the OCR model from being widely distributed during this time.

- ***1980s: Decline of Physical OCRs, Rise of Conceptual "Openness"***

During the 1980s, most open-plan classrooms changed back to traditional, enclosed layouts because of increasing management complexities. Nevertheless, the core idea of openness still prevailed, with its emphasis moving beyond the mere change of the space to the wider scope of educational openness. Teachers have been raving about the idea of developing open curricula, encouraging learners to study at their own pace, and using student-centered methods, which all focus on the learner's experience rather than the limitations of the physical environment. This shift in ideology was a turning point for later understanding of "openness", especially in the digitized world.

- ***1990s: The Digital Shift and Early Online Learning Integration***

With the introduction of personal computing and the internet in the 1990s, the OCR idea grew greatly beyond its original boundaries, which were defined by its physical space. Distance education programs and early online learning platforms enabled students to interact with course content beyond the walls of traditional institutions. The Open University in the UK, along with other similar initiatives, was instrumental in demonstrating that digital technology could be a powerful tool for accessibility and flexibility—two of the major characteristics of the OCR philosophy.

Moreover, this period marked the beginning of debates about OERs that would become the basis for large-scale sharing of open content among various learning communities.

- ***2000s: The Era of Open Educational Resources (OER)***

The decision to make the MIT OpenCourseWare project [30] publicly available in 2001 was a major turning point in the educational landscape, as it allowed anyone free online access to university-level courses with their respective materials.

The Cape Town Open Education Declaration of 2007 once again signified the global movement that advocates openness in education. The OCR Initiative was beginning to change its function, which no longer referred only to the characteristics of the spatially flexible class but also to the manifestations of the principles of open access to educational content, open participation, and global collaboration.

Over time, digital platforms and LMS have increasingly combined face-to-face and online modalities, leading to “open classrooms” that merge traditional and digital learning experiences.

- ***2010s: The Expansion of MOOCs and Social Learning***

The growth of MOOCs during the 2010s largely influenced the openness of the educational system to a great extent [31].

Top educational providers such as Coursera, edX, Udacity, and FutureLearn transformed the concept of educational availability by democratizing quality educational resources while increasing their reach to an audience of practically infinite size worldwide that could now access them at any time. One of the major trends of this era was virtual, interconnected classrooms, where students from different countries, cultures, and academic disciplines collaborated on joint projects. Furthermore, besides traditional OCRs, there were also informal ones, which were social media platforms such as Facebook groups, YouTube channels, and Reddit forums, where peer-to-peer learning and dialogue could be easily facilitated. Initially, the OCR Initiative concept had evolved into a global educational philosophy that promoted cooperation among the different continents, schools, and teachers, as well as embraced inclusion and shared resources across the educational landscape.

- ***2020s: AI-, IoT-, and Data-Driven OCRs***

The COVID-19 pandemic caused a massive and immediate adoption of remote learning models that schools worldwide had to accept open, flexible frameworks almost instantaneously. Today, the OCR Initiative is leveraging various advanced technologies to enhance and expand the educational curriculum, including AI, IoT, and adaptive learning systems. AI-driven tutors and chatbots offer instant help and support, whereas IoT gadgets, such as smart-boards and connected sensors, ease the transition from a traditional learning environment to a virtual one easily and quickly.

By these means, the initiative now epitomizes the essence of openness in all respects – access, technology, pedagogy, and inclusivity, thus serving as the fertile ground for the next generation of smart classrooms with learners' diverse requirements around the globe being taken into account, gradually shaping.

3.2. Components of the OCR

The OCR model relies on three main components: flexible spaces, student-centered pedagogy, and the integration of technology. Individually, these elements transform the learning environment. Still, when combined, they create dynamic ecosystems that enable learners to take an active role in their educational journeys, which is what they do.

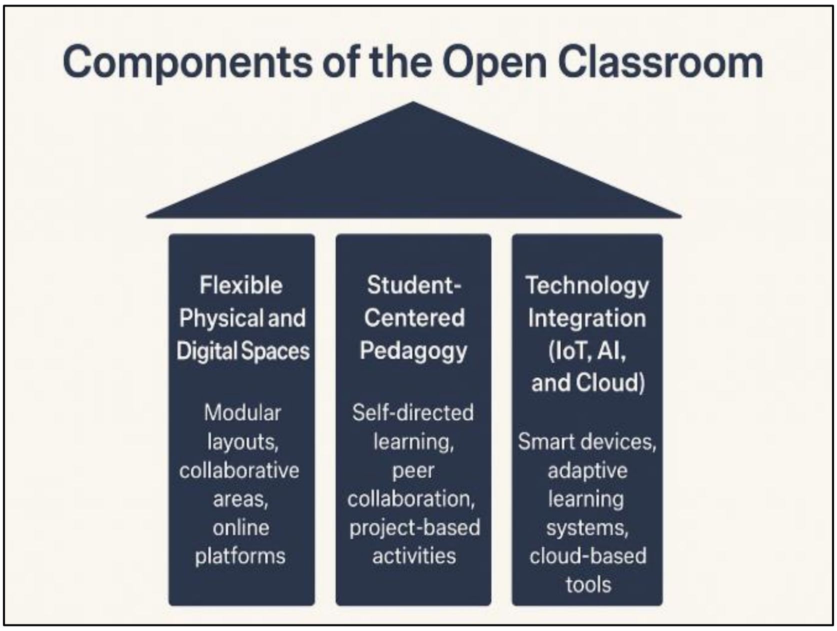

Figure 1-3: OCR Components

- ***Flexible Physical and Digital Spaces***

OCRs give physical and digital spaces equal importance in terms of adaptability. Traditional, rigid classroom and desk arrangements are replaced with modular furniture that can be personalized and even altered by the students themselves to create a social hub, a circle for discussions, and an area specifically set aside for their group projects or individual study. Gradually, this pattern of interaction becomes the norm, while different participants also get the chance to follow their learning style and requirements. In online environments, the concept of flexibility is realized by using sophisticated e-learning platforms, cloud-based collaborative tools, and virtual classrooms. Such facilities allow students to interact with their peers and professors at any time and from any place, thus facilitating the notion of global connectivity. As a result, this method of flexibility ensures that the setting of the learning environment can be adjusted to the needs of various educational activities and objectives.

- ***Student-Centered Pedagogy***

The OCR is student-centered learning at its best. Student-led learning is the focus of this educational design; however, students are still under the teacher's guidance. The mode of learning becomes more pronounced, and students use self-directed exploration, collaborative learning, and project-based activities. These activities, under the guidance of the teacher, lead students to engage deeply with the content.

In this environment, the teacher's role is not that of an authority but a facilitator or mentor; hence, the transition is made from traditional positional power to the new role of a guide. The teacher, therefore, should be a very supportive one who encourages, guides, and provides learners with the tools they need to develop their own ways of handling the curriculum.

This approach not only invigorates and refines creativity and critical thinking skills among students but is also instrumental in instilling in them a feeling of ownership of their learning processes. They are enabled to learn at their own leisure, explore the subjects that interest them, and, in doing so, engage with the broader educational goals in a way that makes sense to them.

- ***Technology Integration (IoT, AI, and Cloud)***

Technology is the backbone of OCR and the main factor that enables the scalability and personalization of the learning ecosystem. The implementation of Internet of Things (IoT) devices such as interactive smart-boards, tablets, and sensors etc., has made it possible for those very spaces to be in direct contact with each other and also to provide instant feedback. Apart from engaging students, these gadgets present important data that is used to guide teaching practices.

Additionally, AI is a pillar in this entire development as it brings along adaptive learning paths and smart tutoring systems that can match the individual requirements of each student. This type of technology makes it possible for teachers to acquire the necessary insights through student performance data and use them to evaluate the strengths and weaknesses of learners. On the other hand, cloud computing has been a primary factor in ensuring the level of accessibility and collaboration that we have at present. It enables anyone from anywhere in the world to have easy access to resources, tools, and communication channels. The perfect union of all these technologies is a magnificent and innovative connected learning ecosystem that surpasses the boundaries of traditional classrooms; thus, it becomes a learning environment that, in theory, can continuously enrich all its users.

3.3. Advantages of the OCR Approach

Even though E-Learning systems have revolutionized education by providing scalability, adaptability, and wide access to information, they often can not deliver a creative and student-centered learning process, which students demand. Conversely, the OCR approach not only takes the advantages but also extends them, emphasizing collaboration, interaction, and flexibility. This model does not exist only in the realm of digital platforms but brings together two technological and pedagogical advancements with the aim of creating a more engaging, inclusive, and personalized learning environment.

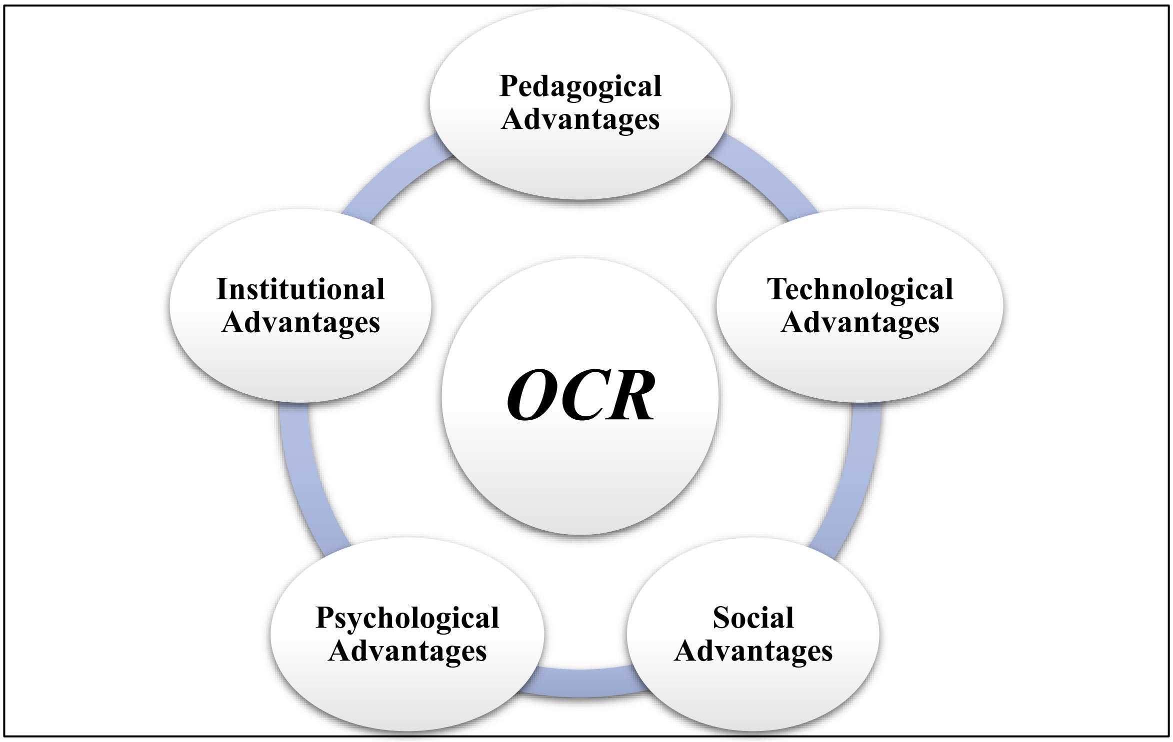

Figure 1-4: The Advantages of OCR

The OCR, by promoting students' participation, providing critical thinking skills, and facilitating the teaching of various learning styles, can resolve the limitations in traditional and pure online education. The following points highlight the major features that empower the OCR to become one of the most successful means of teaching and learning in the future.

3.3.1. Pedagogical Advantages

- Personalized Learning: The OCR design empowers every learner to advance at their own speed, choosing resources and tasks that match their individual learning patterns. Such personalization induces not only a deeper comprehension and longer retention of the learnt because the students get to interact with material consistent with their interests and strengths.

- Student-Centered Approach: The approach brings traditional education to a higher level by diverting the focus from teacher instruction to student-driven research and discovery. Pupils become responsible for their own learning and assume the role, and the critical thinking and independence skills are developed as they investigate the topics that captivate their curiosity.

- Active Engagement: OCRs give preference to project-based learning, experiments, and problem-solving. Students, as active members, get involved in the practical activities that not only deepen their conceptual understanding but also broaden their skills through the application of the knowledge in the real world.

3.3.2. Technological Advantages

- Integration of Digital Tools: The implementation of several digital tools like smartboards, IoT devices, and AI technologies has a considerable beneficial effect on the learning environment. These tools enable the delivery of engaging lessons and the instant provision of the required information from the vast database.

- Access to Global Resources: An OCR surpasses the boundaries of the local area by accessing online databases, MOOCs, and virtual labs. Such global availability broadens students' knowledge and makes their learning journeys richer as they interact with diverse experiences and experts.

- Adaptive Systems: The AI-based algorithm tools allow teachers to give instant feedback and tailor-made suggestions to students. These adaptive systems not only track but also recognize students' progress and areas for improvement, thus ensuring that learners receive the assistance they need to accomplish their objectives.

3.3.3. Social Advantages

- Collaborative Learning: Open educational resources encourage a collaborative spirit in the students, and they are now more successful in involving peer-to-peer communication and co-curricular activities not only with other students but also with teachers. Besides learning interpersonal skills through this common area, students also have the opportunity to experience a sense of community through the collaborative spirit.

- Inclusivity: The adaptable nature of OCRs is suitable for all kinds of learners, as the content is customized to meet the needs and backgrounds of different learners. Such a level of community feeling helps students recognize their worth and supports them along their educational journey.

3.3.4. Psychological Advantages

-Motivation & Autonomy: Students will be more motivated and feel that they have more control over their learning when the OCR composes their learning paths. This decision of theirs empowers them; as a result, they become more interested and take pleasure in their learning.

- Creativity & Innovation: The adaptable, open areas of these classes not only motivate students to delve deeper into their concepts but also facilitate the idea of them being unrestricted by the usual classroom settings. Such independence gives rise to and stimulates the students' creativity and innovation, as they can think beyond the conventional ways and come up with original solutions.

- Reduced Anxiety: The modifiable surroundings of OCRs are instrumental in lowering the stress and anxiety that are usually linked with the occurrence of high-stakes testing and the existence of stringent structures. Children grow strong in places that place a high value on their psychological welfare and offer an encouraging educational encounter.

3.3.5. Institutional Advantages

- Scalability: OCR principles can be practically implemented not only in different classrooms but also in schools or even a country's educational system. Such a wide range of educational settings where this approach can be applied makes it a simple but effective solution for improving educational outcomes.

- Curriculum Flexibility: The open curriculum is more dynamic and can quickly be updated with the latest developments and challenges in the field of education. Such a characteristic of the curriculum ensures that students receive the most relevant and timely education.

- Cost Efficiency in the Long Run: The OCR can represent a substantial saving in costs over time after the implementation of digital infrastructure. Digital content is the resource that has the potential to be shared and reused, which in turn lowers the costs of printing and distributing traditional materials, thus facilitating a sustainable educational model [32].

The OCR method is a significant development in the field of E-Learning. This new model overcomes typical issues that are present in traditional digital learning platforms, for example, the lack of flexibility and interaction. Typically, in closed systems, the learning process is somewhat restricted by the predetermined path that the user has to follow. However, the OCR is all about promoting adaptability, integration, and customization. Such a revolutionary concept provides the users with the possibility to experimentally create their understanding of the topics taught in a lively and customizable space.

3.4. Challenges of the OCR Approach

Although the OCR concept has major advantages that include adaptability, personalization, and inclusivity, its application is still fraught with some issues. Its features that distinguish it from others also contribute to its intricacy. Unlike the conventional systems that emphasize structure and uniformity as the key elements, the OCR demands a lighter, more flexible, and interactive approach, which includes the active participation of both learners and educators and their continuous adjustment. This change of paradigm keeps the door open to new problems related to student motivation, ensuring that resources are accessed on a fair basis, integrating different kinds of technology, and assessing in ways that are different from the standard ones.

Besides that, the openness of the model can lead to different interpretations of the learners: they may become confused about what to do by the multitude of options available; the instructors may find it hard to strike a balance between complete freedom and guidance; and the institutions may have problems in keeping the necessary technology and administration. Such challenges exemplify the need for robust mechanisms in place that can simultaneously safeguard the advantages of an OCR and address its vulnerabilities.

- ***Information Overload and Resource Fragmentation***

Open educational environments usually rely heavily on digital content libraries. These include different materials such as textbooks, recorded lectures, multimedia presentations, and interactive simulations. In addition, the vast array of resources can significantly benefit education; however, they can also confuse both the teachers and the students. It is quite possible that students become overwhelmed with the abundance of data presented in a disordered manner and, therefore, have difficulty finding the appropriate materials that suit their specific needs. At the same time, educators might struggle with designing proper learning pathways due to the distribution of resources, causing difficulties in forming a teaching storyline.

- ***Lack of Real-Time Adaptivity***

A primary issue with traditional e-learning systems inherited by OCRs is that they are unchangeable and, therefore, are not capable of meeting the changing needs of users in real-time. Differences in the learning pace, prior knowledge, and unique learning styles are quite rarely discussed. Consequently, some students may find they are at a disadvantage because they have not been given enough help, while others may lose interest because the content is not challenging enough for them or moves too slowly. The absence of a responsive adaptation to the student's needs defeats the personalized learning experiences that OCRs aim to provide.

- ***Limited Student Engagement Monitoring***

Another main issue with OCR environments is keeping students actively involved in the lesson. A major difference between open and traditional classrooms is that the former do not have physical or psychological indicators that can be used for control of students' attention and motivation levels. Teachers often encounter situations where they don't have the tools or strategies to identify disengagement, distractions, or shifts in concentration among students. The invisibility of these signals makes it difficult to promptly support struggling students, which can lead to declines in academic performance and interest, making progress almost impossible.

- ***Assessment and Progress Tracking Difficulties***

Assessment of education results in OCR remains a difficult issue. Usual standardized assessments may fall short of reflecting the detailed learning of students, especially in varied and self-paced environments. Besides, the use of feedback based on self-reports can produce data that are not quite reliable due to bias or misunderstanding of one's progress. The lack of an accurate system for monitoring both personal and group growth makes teachers lose the most valuable information on their students' learning process, thereby making them less likely to provide the right support and guidance.

- ***Scalability and Resource Management***

OCRs, by their very nature, aim to serve larger and more diverse learner populations; nevertheless, the process of scaling up such settings is still fraught with numerous logistical difficulties. The handling of massive amounts of learner data, maintaining the quality of courses, and ensuring a uniform educational experience in different settings are some of the challenges that can hinder infrastructures to a large extent. Furthermore, the extensive scaling of these systems often results in the loss of personalization, which empowers a mere standardization of instruction and disengagement of learners.

- ***Equity and Accessibility Concerns***

Although OCRs are designed to decentralize education, there are still major differences in access to technology, internet connection, and digital literacy. Children who live in underprivileged areas, those who have little access to technological resources, or individuals with disabilities may encounter such obstacles that they will not be able to take part in open learning activities. Achieving absolute inclusiveness in such educational environments is essential, as it is the key to fulfilling the promise of open learning.

- ***Instructor Workload and Pedagogical Adjustment***

Teachers in OCRs face the double challenge of adjusting to new methods and managing more complicated teaching environments. Besides content preparations, teachers need to supervise very different types of learners, respond to their differing needs, and keep them all engaged. Such a rise in teacher duties can easily result in the loss of teachers' enthusiasm and motivation if they are not provided with sufficient support in their multifaceted roles.

- ***Trust, Data Privacy, and Ethical Issues***

OCRs' very existence depends largely on the use of various digital tools and the gathering of learner data. Nevertheless, suspicions about the safeguarding of data, the moral use of users' private information, and the level of openness in the functioning of automated systems can constitute severe obstacles to the implementation of such models on a large scale. It can be expected that students, parents, and educators will hesitate to adopt OCR systems when there is not enough trust and ethical guarantees that they will not be disappointed in their beliefs, and, as a result, ultimately impeding the progress and acceptance of this teaching method.

Understanding these challenges facilitates a better understanding of the underlying issues blocking the way to the OCR approach that is both sustainable and effective. This important reflection acknowledges that, practically speaking, the model has limitations, and simultaneously clears the way for future inventions that aim to restore and improve the educational experience.

4. The Role of AI, Deep Learning, and IoT in the OCR

The OCR method revolutionized the educational system, highlighting the importance of the learning environment's flexibility and adaptability. Nevertheless, the real possibility of the OCR remains only when it is combined with state-of-the-art technologies, such as AI, deep learning (DL) [33], and IoT [34]. Such technologies combine to transform the OCR, a delightful idea from history that grew into a fascinating, intelligent, and highly personalized learning environment.

The application of AI and DL to the education sector has resulted in the emergence of a flexible educational process. Fundamentally, the use of these systems is to handle complex and large-scale data mining activities, identifying and illustrating behavioral patterns of the students. One instance of this is the simple use of predictive analytics to foresee possible learning difficulties, allowing educators to intervene proactively before students face significant challenges difficult to solve alone. DL models are highly effective with complex and unstructured data [35], such as speech recognition, handwriting, and video interactions. This feature is essential for identifying student participation and focus levels, thus making it easier to understand the learners' individual educational experiences. Furthermore, virtual assistants powered by AI also make a great contribution to the vitality of the learning environment, as they make it possible for students to easily access a large number of resources, ask questions, and get immediate answers. Students are allowed to guide their own learning process through on-demand assistance, which fosters their autonomy and motivation.

Another significant aspect is IoT and its role in tearing down the educational walls by linking the physical and the digital parts of a class [36]. The employment of smart sensors, wearable devices, and interconnected educational tools has made it possible to measure, for example, students' attention spans, cooperative engagement, and even feelings. By using real-time data streams, IoT upgrades the educational experience, providing teachers with a substantial part of students' wellness and learning dynamics. In a genuinely free classroom, the incorporation of IoT would mean that pedagogy is no longer confined to the digital screens but is instead enriched by the endless, real-world interactions that can always be recorded and decoded. This all-round approach not only facilitates participation and accessibility of the classroom but also its adaptability to the different needs of learners.

When these three technologies are combined, AI, DL, and IoT provide a solid technological foundation for an innovative, next-generation OCR. Such a classroom will definitely make the student the focus of the whole data-driven and adaptable to the instructional needs of both individual and group learners, not only learner-centered but also. This strong and fruitful combination of modern and progressive teaching methods and high-tech tools opens the thesis's new chapters, where the focus shifts from the theoretical basis of the OCR to the practical aspects of designing, implementing, and evaluating intelligent educational models that make these concepts real-world applications.

5. Conclusion

The chapter laid the foundation of the thesis by studying E-Learning systems and the OCR approach. We began by exploring E-Learning evolution through various stages and its core components with real-life cases. We examined the advantages of the technology, including access, scalability, and personalization of the learning process, together with the inconveniences that involved inflexibility and limited adaptability of the system.

This chapter has examined the basic concepts in an exhaustive manner, which are the main ideas behind the OCR approach. We follow its developmental timeline, from its beginnings with the ideas of progressive education to its present-day usage in various educational settings. We not only recognized but also extensively discussed these three major areas as the cornerstone for OCR: flexible learning atmosphere, student agency, and collaborative opportunities. The evaluation has demonstrated the advantages of open classes, including increased student participation, individualized learning experience, and the development of a community of learners. Despite this, it has also raised issues about the prominent obstacles in the ambit of structural and organizational problems, as well as concerns about how to implement this kind of education, which, by nature, is diverse, and how to keep students continuously engaged not only now but also in the future.

We recognized the intricacies of these issues and highlighted the transformative power of new technologies, especially the use of AI, DL, and IoT. These technologies are instrumental in the development of the OCR from only a conceptual model into a smart, data-driven, and adaptive learning system, which can adjust to the needs of both teachers and students.

By clearly defining the basis of this understanding, the chapter acts as a bridge to the next phase of the paper: an in-depth investigation of the related academic works. The next chapter will be a detailed study of the main research that has addressed similar problems of e-learning and OCR environments. Among the various topics covered in the review, particular emphasis will be laid on innovations like AI-driven search engines that make learning more personalized through content delivery, the employment of virtual assistants for student support, the installation of sensors for the real-time detection of student engagement levels, and the development of predictive models for student progress. This review will acknowledge the academic context of our current research while identifying the gaps in the research field that provide the motivation for our proposed contributions.

Chapter 2: Related Work

1. Introduction

In the previous chapter, we explored eLearning and the Open Classroom Initiative from the ground up, figuratively dissecting their conceptual frameworks, technological underpinnings, and pedagogical implications. While these paradigms have largely democratized education access beyond traditional settings, our evaluation revealed the continued occurrence of certain problems, such as the static nature of educational content, the difficulty of sustaining learner engagement, the lack of sufficient personalization of learning paths, and poor knowledge retrieval performance in open educational ecosystems.

The research community has taken several measures in response to the mentioned restrictions to improve the situation. They have introduced numerous innovations, including, among others, the implementation of intelligent search algorithms that facilitate content discovery, virtual assistants that enable easy and adaptive interactions between users and systems, student engagement detection tools that use computer vision and deep learning, and predictive modeling for monitoring learner progress that guides personalized learning strategies. These developments together provide the background of the technology and methods that we use in our study. This chapter summarizes those relevant works in detail and highlights the main methodologies, models, and techniques that have been used in research for the areas identified.

2. Related Work

2.1. Works Related to prediction and search engine

The use of smart search engines has risen significantly in digital learning platforms. It is often the case that traditional search mechanisms, based only on keywords, simply fail. One of the major reasons for this is their lack of user intent recognition. Due to this restriction, the effectiveness of these guidance systems in providing meaningful assistance to learners is severely limited. Considering these difficulties, newer improvements have led to the development of semantic search engines. Such engines rely on Natural Language Processing (NLP) techniques and complex Deep Learning models, for example, BERT (Bidirectional

Encoder Representations from Transformers) [37] and GPT (Generative Pre-trained Transformer) [38], to name just a few. However, despite such remarkable achievements, their deployment in closed educational settings is still very patchy and largely untapped.

2.1.1. Artificial Neural Network (ANN)

An ANN is a computational model inspired by the structure and function of the human brain [39]. It consists of connected nodes, generally known as neurons, that are grouped in layers. Normally, an ANN is composed of three major types of layers: the input layer, the output layer, and one or more hidden layers. In the training phase, the ANN takes input data, processes it through the different layers. Each pair of connected neurons has a specific weight, which shows the power and significance of the link in question. When the network is given data, it changes these weights, which is a process known as backpropagation; thus, the network can learn and improve its forecasts.

Guettala et al. [40] examined how artificial intelligence with a generative core could revolutionize the educational sphere, particularly by facilitating adaptive and personalized learning. Their study presents a comprehensive roadmap of the possible integration of generative AI into the adaptive learning landscapes. It considers various types of recommendation systems and also provides insights on how one-to-one educational content can be sufficiently designed to cater to the needs of individual learners. The authors point out the possible benefits of generative AI tech to the development of more flexible and individualized learning experiences. Mahmood et al. presented an extensive study on multimedia data retrieval, developing a comprehensive framework that uses neural networks of Multilayer Perceptrons (MLPs), a type of artificial neural network, for effectively retrieving and classifying a wide range of content including text, images, and audio [41], [42]. Their innovative approach combines traditional text-based retrieval methods, such as Latent Semantic Indexing (LSI) and Vector Space Model (VSM), with a deep learning component specifically designed to handle feature-extracted data (e.g., data obtained through wavelet transform), thereby enhancing the retrieval process.

In this research, the authors have used Artificial Neural Networks to automatically assign categories to the web pages that are pre-defined, such as education, news, job search, and many others. Such automation has enabled retrieval to be made to a much larger extent from the metadata and has allowed a more user-friendly search navigation [43]. The described system made use of sophisticated methods for achieving the automated extraction of features from web page content and relied on backpropagation algorithms for the correct categorization, thus simplifying the whole information retrieval system and enhancing the user's experience in finding the required multimedia resources [44]. Balasamy and Athiyappagounder designed an e-learning recommender system utilizing a four-layer MLP to anticipate and suggest appropriate learning materials [45]. Their deep neural network (DNN)-based method outperformed the logistic regression baselines.

2.1.2. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a sophisticated feedforward neural architecture that consists of interconnected artificial neurons [46]. Such networks are especially efficient for handling big image data, thus their widespread use in the area of computer vision [47]. CNNs possess various characteristics, including trainable weights, bias terms, and the notion of parameter sharing via convolutional filters. The design of CNNs allows them to be very successful in gaining spatial hierarchies in visual data; thus, they can be used in facial analysis and attention tracking with great precision and speed [48], [49].

The authors in [50] explore a novel method of fusion between Convolutional Neural Networks (CNNs) and graph analysis techniques. This merger aims to straighten the relevance of course recommendations in e-learning setups. The research team, by employing both techniques, aims to give the learners more accurate and situationally relevant recommendations, thereby enhancing user engagement and the overall learning experience.

One study by Nannini et al. [51] led to the creation of a CBIR web search engine, which is smartly designed to combine the features of ResNet50 and ResNet50v2 as feature extractors with Locality Sensitive Hashing (LSH). The merger enabled similarity searches and best results indexing, which are tested with mean Average Precision (mAP) benchmarks. These developments are essential for providing images that most closely match the users' queries effectively, hence improving the user experience in image search applications. In another study, they disseminated one that dealt with the functionality of VGG16 and MobileNet in CNN-based CBIR systems [52]. Essentially, the question was how these models could be made efficient, together with indexing libraries such as Faiss and Annoy, for handling the balance between retrieval accuracy and response time. Their findings underline the potential for improved performance in scenarios that require both quick access and trustworthy image retrieval.

Moreover, Putzu et al. developed a content-based image search (CBIR) system, which relies on convolutional neural networks (CNNs) not only for feature extraction but also for conducting relevance feedback loops [53]. The system upgrades its function by adjusting the CNNs through examples provided by the user of both relevant and non-relevant images. With repeated interactions, the system strives to reduce the semantic gap between what users expect and the retrieval accuracy, thus making image searches more efficient all the time. Additionally, Sarasu et al. introduced a novel and unusual model in the area, that is, the Semantic Featured CNN (SF-CNN), which is specifically designed for text document retrieval only [54]. SF-CNN implements a CNN-based architecture that facilitates the extraction of semantic features as well as the resolution of the ambiguity of the terms that occur both in the queries and in the documents. This system led to almost 94% higher accuracy than traditional baseline algorithms, which was its main feature.

2.1.3. Deep Neural Networks (DNN)

Deep Learning architecture is a more complex and specialized model of Artificial Neural Networks (ANN) that attempts to raise the prediction accuracy by having more than one hidden layer in the network [55]. Such a multi-layer configuration allows the network to uncover complex patterns and intricate relationships in the data. Whereas the traditional neural networks usually have only one or two hidden layers, Deep Learning has much more hidden layers. This extended design elevates the machine's potential to do feature extraction and representation learning and thus makes it possible to work with more complex datasets and tasks.

Chen et al proposed a DNN-based system that intends to provide non-verbal communication cues to presenters, in open classes, in real-time [56]. The processing of facial expressions, gestures, and body posture, through a deep model pre-trained with presentation videos, is the core of the system. The real-time system is, however, in training mode. It gets feedback from audience members' devices (e.g., via an app or sensors), thus allowing the presenter to adapt and improve performance dynamically. Consequently, the teacher becomes more engaging and interactive, which facilitates the successful delivery of the open classroom method. This work [57] introduces a DNN-based vision-monitored smart-classroom system that operates under an osmotic IoT framework. The design is meant to identify the physical features of the classroom, for instance, students, writing, and learning objects, through a camera feed. The DNN was assigned to the task because of its excellent image recognition abilities and its power to keep performance even when a slow network and low bandwidth, which are frequent in real-time class situations, occur. The study is basically a kind of endorsement of DNN technology as a sturdy AI teaching system working at the edge where latency-sensitive decision-making is needed.

Developers present a recommendation system for the learning path, which can be expanded and makes use of Graph Neural Networks (GNNs) along with reinforcement learning methods [58], [59]. Such a combination allows the system to reconfigure the order of learning activities according to the user interactions and likes for an indefinite period. Consequently, an adaptive learning setting that modifies itself to optimize the educational progression of every learner is formed. The authors of this paper have come up with a context-sensitive deep learning setup that aims to offer on-the-fly changes of user preferences for custom-tailored suggestions [60]. The Deep Adaptive Interest Network (DAIN) uses cutting-edge AI skills to grasp and react to users' changing likes, thus making the recommendations more practical and efficient in the educational field.

Dash and Kumar unveiled a non-verbally gesture-based document prediction system, which employs a hybrid CNN-GRU model to facilitate non-verbal communication in schools [61]. To put it simply, CNN handles the location feature extraction of the gesture images, whereas the GRU captures the changes over time for the gesture sequences. The system made it possible for users like teachers or students to access documents or carry out activities via shortcuts and hand gestures, thus creating a touch-free and user-friendly interface. It is the new way of multimodal search, where gestures are the queries. The model was very successful in achieving high real-time accuracy, and it was better than standalone CNN and LSTM models; thus, it made deep learning's potential very strong for intelligent and context-aware content retrieval in smart classrooms.

2.1.4. Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are one of the crucial categories of artificial neural networks that are especially tailor-made for handling sequential or time-series data [62]. RNNs, as a distinguishing characteristic, keep the context of the previous inputs through their internal memory. This feature enables RNNs to apply past information to the present inputs, which consequently determines not only the present outputs but also the future forecasts [63]. Unlike CNNs that take in only spatial data and depend heavily on convolutional layers for feature extraction, RNNs are very good at handling time-dependent data sequences.

Therefore, they are almost perfect for such tasks as language modeling, speech recognition [64], and predictive analytics. The authors in [65] presented a user-centric course recommendation system for e-learning platforms. The system, which relies on BERT and complex deep learning algorithms, personalizes course suggestions by analyzing user data, thus facilitating and optimizing online education for every user.

Ahmad et al. have presented a novel two-level RNN model that can efficiently grasp the search context at the query and task levels for multi-query sessions [66]. The new model enables both real-time document ranking and the utilization of query suggestions through the use of recurrent attention layers. These layers, which uncover complex relations between consecutive user search tasks, thus making significant improvements in search relevance and user satisfaction metrics, have found their application. Hidasi et al. [67] designed the GRU4Rec model, which is one of the first leading examples of using GRUs to capture item-to-item transitions in short user sessions, particularly in domains such as video consumption. The new model not only outperformed conventional item-based collaborative filtering methods across a range of crucial metrics, such as recall@20 and mean reciprocal rank (MRR), but also enabled accurate next-item predictions in real-time, even when the recommendation scenarios are changing rapidly.

Sangamithra et al. developed an innovative, intricate hybrid model that merges RNN and LSTM networks for search re-ranking tasks [68]. This algorithm utilized all user feedback, including comments and prolonged search history, to improve the ranking of search results. The extensive experiments conducted on datasets taken from leading search engines Bing and DuckDuckGo have shown not only significantly high-performance measures of precision and recall but also, surpassing traditional information retrieval models. Ge et al. (2019) presented a hierarchical RNN design that depicts the user's sequential query sessions [69]. With the help of query-aware attention, the RNN selects relevant contextual user profiles for a more accurate and personalized search result. Tested with authentic commercial search logs, the method granted a considerable jump in ranking correctness over traditional personalization baselines.

2.1.5. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a stochastic optimization technique that takes inspiration from the social behaviors of animal swarms, like bird flocks or fish schools [70]. The function of this method is basically collective intelligence, where a group of agents, particles, move about a given search space to pick out the best solutions. A particle corresponds to one of the possible solutions, and it changes its location depending on its own experiences as well as the experiences of the particles in the vicinity. PSO achieves this by social interaction simulation to find the best way within a complicated area for optimization to get the best solution. The utilization of this method has spanned widely across different areas, for example, tech, the stock market, and machine learning [71].

In [72], the authors introduced a new hybrid semantic clustering algorithm that combines Particle Swarm Optimization (PSO) with the k-Means clustering method seamlessly to organize documents systematically. This method is, fundamentally, designed to enhance the clustering stage by using a semantic similarity metric that involves an ontology, thus depicting the relations between the documents at a deeper level. Here, PSO is used to facilitate the cluster centers' movements in the vector space, which are the points that show the semantics of the annotated source of knowledge. The purpose of this optimization is to enhance the semantic coherence of clusters and, as a result, improve their overall quality. Through the improved indexing of documents, the system can increase the efficiency of the retrieval as well as the general search function.

Moreover, Ramya also explained an elaborate PSO technique with the main purpose of improving web document retrieval [73]. In this case, PSO was effectively applied to achieve the best performance in both indexing and similarity matching, which in turn brought outstanding outcomes. This approach, when tested on both small datasets such as the CACM database and large datasets like RCV1, not only revealed considerable progress in retrieval speed but also the relevance of the results.

Ramya and Shreedhara have taken the initiative to extend the earlier works by Ramya to put forward a PSO-based document retrieval plan that integrates a newly designed similarity metric called SMDR (Similarity Metric for Document Retrieval) [74]. Their method aims to improve retrieval by employing the search capabilities of PSO to achieve a higher level of precision and a lower response time. The improvement was particularly visible in the performance test using the CACM corpus, which indicated that their method had advantages over traditional IR models. The progression to these levels of performance highlights the effectiveness of exploiting PSO-based semantic clustering approaches for the handling of documents in different scenarios.

2.1.6. Summary table for Prediction and Search Engines

Table 2-1 presents a comprehensive summary of the related works. The table displays the advantages and disadvantages of each mentioned method, providing a clear comparison for reference.

Table 2-1: Comparison Between the Chosen Methods for Prediction and Search Engines

Method	Advantages	Disadvantages	Performance Insights	Benefits of Hybrid PSO/ LSTM
ANN [40], , [44], [45].	- Ensuring that data is preserved across the whole network. - Possessing the capability to work with sensitive information. - Featuring a distributed memory system.	- Reliance on hardware components. - Inexplicable behavior within the grid. - Difficulties in illustrating the issue to the network.	- Suitable for basic text classification tasks; however, due to limitations in processing sequential and contextual information, they are seldom the optimal choice for intricate text prediction challenges.	- Designed specifically for sequences, memory techniques effectively manage dependencies across lengthy texts, leading to improved performance in text prediction.
CNN [50],[51] [52],[53] , [54].	- Characteristic extraction is automated. - Image recognition and classification exhibit high precision. - Computational needs are minimized.	- A significant amount of labeled data is required. - There is limited effectiveness for sequential data. - Processing speeds are slower. - The training process is time-consuming.	- It is best for capturing local text features and is applicable for text classification as well as feature extraction.	- Contextual information and long-range dependencies in sequences can be captured.
DNN [56], [57], [58], [60], [61].	- Achieves high accuracy scores. - Widely used across all areas of deep learning.	- Requires substantial data to prevent overfitting and achieve optimal performance. - Incur substantial computational costs. - Training poses challenges, such as underfitting or overfitting.	- DNNs are suitable for text-related tasks that do not involve sequence modeling, like classification.	- They retain sequence information.
RNN [65], [66], [67], [68], [69].	- Can handle inputs of any length. - Using models that retain information over time is beneficial, especially for time	- Using activation functions like ReLU or tanh makes processing long sequences challenging.	- Appropriate for text prediction tasks where context is essential, though they struggle with long-term dependencies.	- Designed to manage long-term dependencies more efficiently, making them preferable for complex sequence

	series predictions. - Even with larger input sizes, the model size remains constant. - Weights can be shared across different time steps.	- The computational process slows down due to repeated patterns. - Training RNN models can be difficult. - Issues such as exploding and vanishing gradients may arise.		tasks.
LSTM	- Superior at addressing long-term dependencies. - Highly effective in modeling intricate sequential data. - Less prone to the vanishing gradient issue.	- Requires greater computational resources and time. - Vulnerable to overfitting when dealing with smaller datasets. - Large datasets are essential.	- Very potent for text prediction, offering a strong method for modeling sequential dependencies and long-term context.	- PSO can be utilized for optimizing hyperparameters or model weights, enhancing the text prediction capabilities of LSTM.
PSO [72], [73]**.** [74]**.**	- Unaffected by the scaling of design variables. - Easily parallelized for simultaneous processing - Possesses very few algorithm parameters. - Acts as a highly efficient global search algorithm.	- Prone to getting stuck in local optima in high-dimensional spaces.	/	/
Hybrid (PSO/LSTM) Proposed	- Through PSO, tuning hyperparameters significantly boosts the predictive accuracy of LSTM. - Selecting suitable regularization and dropout parameters improves generalization. - Automated optimization decreases the time required for manual tuning. - This method is flexible and applicable to various deep-learning tasks and models.	- High computational expense. - Over-optimizing hyperparameters for a specific validation set can lead to poor performance on unseen data.	- By harnessing PSO's capacity to systematically explore and fine-tune hyperparameters, LSTMs can enhance their performance, making them more adept at understanding and forecasting text sequences.	/

2.2. *Works related to Virtual Assistants*

Several research works have extensively investigated the use of AI-powered virtual helpers in facilitating students' learning journey. The AI assistants vary greatly in complexity, for instance, simple rule-based chatbots, which operate according to a set of predefined scripts, and highly developed Transformer-based NLP models that can generate detailed responses for an almost limitless variety of learner questions. The well-known instances of this category are the platforms like Google Assistant and Amazon Alexa, which give general support, in addition to educational chatbots like EdSights, a specialist that focuses on the educational context. Such tools help users in different ways, like learning their way through platforms, answering FAQs, managing courses, and giving reminders for assignments and deadlines on time [75](Kukulska-Hulme, 2020). Additionally, the adoption of these AI-led assistants is becoming a key factor for the future of interactive and personalized learning practice, which is the ultimate goal of student engagement and outcomes.

2.2.1. Transformer-based Virtual Assistants

Transformer-based virtual assistants have revolutionized the entire model of human-computer interaction, particularly in the educational and professional areas. Such cutting-edge systems rely on multiple deep architectures, including T5, GPT-2, GPT-3, and Flan-T5. In essence, they enable users to have more natural, logical, and contextually appropriate conversations due to their enhanced NLU and generation abilities. The recent study has been very intense, with the focus on combining these strong models into numerous chatbot frameworks in order to get the largest possible improvement in the support given to learners. This combination makes it effortless not only to provide immediate answers to students' questions but also to personalize the educational experience by adjusting it to the individual learning styles and needs. Moreover, these virtual assistants can help educators to have a more efficient workflow by using automated administrative tasks as grading and scheduling, which enables them to focus more on instructional quality.

Moreover, these models have the potential to open new channels for teaching and content creation by automating customized educational materials. For example, by processing educational data and the ways learners interact with it, such systems could come up with quizzes, summaries, and instructional texts that meet curriculum requirements and student capabilities. Basically, the deployment of transformer-driven AI personnel in school settings is a win-win situation, as it can elevate academic achievement and administrative effectiveness.

Baha et al. in their research, combined the use of the CamemBERT model, a French version of BERT, with a chatbot architecture called Xatkit [76]. The encoder retrieved deep contextual representations of student queries while the decoder inferred the intent and produced a reply. This intent recognition was very instrumental in detecting instructional intents such as clarifying by example, requesting extra resources, or making an unfriendly statement. The educational influence of this linkage traversed the spectrum: the instructors' cognitive load was reduced due to the automation of straightforward query handling, learners' autonomy was enhanced as they could now practice self-questioning, and the general classroom responsiveness was elevated; thus, the burnout among educators was kept at a minimum. Technically speaking, CamemBERT embeddings were the means through which semantic encoding was carried out, while the decoder, by all indications, was acting as a classifier of the task-specific kind or a dialogue manager that grouped intents and made queries match with the suitable response formats or content delivery modules. Mathur et al utilized T5 and GPT-3 to generate logically coherent and semantically rich test cases from text-based requirements or academic descriptions [77]. These models were trained to detect conversation motifs (topics) and to automatically produce question formats from the identified keywords and the underlying intent. The educational benefit of this solution was the automation of the exhausting process of creating tests and quizzes. Moreover, the process became easier for instructors because they could use the topics to ensure the alignment of learning outcomes by maintaining fidelity. Besides that, the technique allowed the use of non-technical instructors to produce good assessments written in plain language. On the technical side, T5 was fine-tuned in text-to-text mode to convert educational text into question templates, whereas GPT-3 was used to bring in more linguistic creativity and variety. Some of the post-processing operations may have included grammar checking, terminology matching, and curriculum alignment.

In their work, Bird et al. developed a T5-style chatbot that not only understood the words of the students but grasped the meaning at a semantic level [78]. This helper was not looking for keywords only, but comprehended users' whole phrases and used the context embeddings to identify the gist of the communication and the social aspect of the interlocutors. The educational impact of this method was such that it provided a user-friendly interface, which allowed the interaction to be deeper, as now the emotional state or social hints of the interlocutors could be inferred. Moreover, the virtual assistant could be regarded as a conversational knowledge base that was able to answer any open questions. The idea was that the T5 model encoded the context at the sentence level and then linked it with the correct output for the particular downstream task. Possibly, the chatbot was on a semantic intent-slot filling setup that employed T5 semantic abilities for the understanding task. Sallove et al. presented Vec2T5, an innovative method that combined Word2Vec to create vector representations of user queries for intent recognition, and T5 to output responses. Such a modular design allowed for better performance when dealing with both short commands and long-form questions [79]. The educational influence of this synergy traversed the spectrum between speed, which was the prerogative of Word2Vec, and semantic depth, which was attributed to T5. As a result of this further development, the chatbot's reaction time was reduced, and the relevance of the answers remained high, all with less computational cost. Technically speaking, the pipeline was probably a two-step process: the first step was quicker intent detection using Word2Vec, trained on the specific domain corpus, and then feeding the results to T5's decoder to obtain relevant and personalized responses.

Wang and Issa demonstrated how T5 and GPT-3 could rephrase the content of a lengthy and complex regulatory document, for instance, rules by the Occupational Safety and Health Administration (OSHA), into a set of natural language questions [80]. Their comparison revealed that while T5 was able to produce a greater number of varied and pedagogically useful questions, GPT-3 seemed to be more structurally consistent in its outputs. It made an enormous educational impact, as it not only promulgated complex legal or technical texts as teaching resources but also provided educators with the tools needed to create formative assessments and furthered engagement and understanding by offering automated questioning. On the technical side, T5 required a lot of work to be adapted for the different parts of the long regulatory text. They probably used evaluation metrics for diversity, like BLEU, ROUGE, and question entropy, to assess their results.

This research [81] examined model adaptation by comparing Flan-T5 and Lamini Flan-T5 to standard GPT-3 across different experiments, such as abstractive summarization and conversational AI. One of the major takeaways of this study was that fine-tuning the transformers for dialogues related to the task greatly increased the chatbot's response quality in terms of both continuity and pertinence. The educational aspect was that it allowed the accurate adjustment of chatbots for different educational markets, like biology tutoring or language practice, as well as the improvement of the logical flow of chatbot responses to maintain context throughout several turns. Moreover, this research demonstrated the importance of fine-tuning over pre-trained models; thus, model adaptation benefits for higher educational applications are highlighted.

Yin et al. have developed educational chatbots with metacognitive feedback mechanisms that allow learners in science subjects to reflect on their learning processes and improve comprehension significantly [82]. Allen et al. came up with the Q-Module-Bot, which is a combined generative Q&A system that mixes one type of data retrieval with transformer-based generation to achieve real-time student interaction on structured course content and to facilitate the educator's administrative work [83]. EduChat, a large-scale educational LLM, is a system by Dan et al. that has been pretrained on educational corpora and fine-tuned with instructional design insights. It has the capabilities of Socratic teaching, essay assessment, and providing emotional support to students, teachers, and parents, demonstrating the close pedagogical alignment of LLM capabilities [84].

Unimib Assistant created a Retrieval-Augmented Generation (RAG) chatbot for university students to serve context-specific student needs by combining knowledge-base retrieval with generative dialogue. However, the system also disclosed the limitations of answer accuracy and source transparency [85]. Likewise, Jill Watson [86], being a modular teaching assistant built on ChatGPT, is capable of processing multiple large documents, integrating safely with classroom workflows, and reducing hallucinations by skill-based API modules, a virtual teaching scalable and reproducible infrastructure, as demonstrated. Moreover, university-developed chatbots at the institutional level are trained on question-answer pairs specific to academic administration and campus life, like KatzBot, which has shown better domain relevance and accuracy than standard LLMs [87].

2.2.2. NLP-Based Assistants

Natural Language Processing (NLP) is a significant component of several smart systems, developed to mimic human language comprehension. Such a technology makes possible a communication that is instantly, context-aware and empathetic, all very important requirements in the mentioned fields of application, such as education, health care, customer service, or smart home environments. NLP-powered machines employ several methods, such as syntactic parsing, semantic analysis, and language modeling, which enable them to understand a given text or speech and, in this way, to provide a proper and logical follow-up.

The purpose of creating the Intelligent Tutoring System was to improve online learning environments. It uses NLP technologies to identify key entities and understand the student's intent in a query. The system can thus give detailed explanations or materials on its own, without the need for an educator's direct intervention. Therefore, this method not only lessens the mental strain on faculty members—who can then concentrate on other difficult teaching tasks—but also promotes students' autonomy, as they become more self-regulated and interested in their education.

The comprehensive review by Maity & Deroy is about the application of large language models (LLMs) in Intelligent Tutoring Systems (ITSs) [88]. The paper explores various aspects, including the automation of question generation, the adaptability of conversation systems, the evolution of personalized learner feedback using different educational strategies, the methods implemented for the prevention of bias in data, and the ethical issues from the use of these technologies. This review level of understanding is very valuable to grasp the implementation of various new designs, especially the ones that rely on T5 architecture, in the ever-changing educational technology domain. The Bank Chat Bot is an additional NLP-based technology that was created solely for the financial sector [89]. To handle a customer request, the chatbot uses the NLTK library, which comprises the same preprocessing operations as are discussed in the last paragraphs. After cleaning, the input is again transformed into numbers that a machine learning classifier takes as input and which it outputs as a category of the intent of the inquiry. Such intents can be, e.g., balance inquiries, fund transfers, or loan information. An easy-going banking assistant is the end-use that this technology provides to one who can comprehend multiple user inputs and react with a suitable, already-prepared answer.

To make it more understandable, the IRON Intelligent Personal Assistant (IPA) system was developed in a different field to help disabled people who are users of voice-controlled IoT devices [90]. A user of this system can speak normally or in a complicated way and still be understood by the system, which uses several techniques like tokenization, named entity recognition, and command classification to accommodate the voice inputs. With the help of NLP-based parsing, it figures out the device as well as the operation (e.g., turn on, open, close), giving the user the power to operate the smart home without using their hands. The project is one of many examples of the potential for NLP to overcome the barriers of accessibility. Another is that Lekova et al. created a voice assistant powered by GPT-J, a large open-source language model for human-robot interaction [91]. This assistant, through advanced NLP pipelines and cloud-based speech synthesis, makes a natural conversation with the user with a prompt and accurate response. Designed especially for service robots, the smart assistant provides answers to user questions in real time and sounds like a human voice. So far, the assistant impresses with its interaction skills, as it can remember topics during dialogues, giving an example of combining NLP technology and robotics for a better customer service experience.

Graesser et al designed AutoTutor, a ground-breaking conversational Intelligent Tutoring System (ITS) that aims to improve learning in conceptual physics and computer literacy [92]. The new system features mixed-initiative dialogue, which allows a more interactive and engaging communication between the tutor and the learner. Besides this, it applies latent semantic analysis and speech-act classification as advanced techniques for the efficient decoding of students' inputs and provision of appropriate responses. The use of animated avatars has made the engagement more appealing as it creates a livelier learning atmosphere. The study has found that AutoTutor may result in very strong learning gains, approximately 0.8 standard deviations of student performance improvement can be estimated. PhysicsAssistant is a sophisticated multimodal tutoring robot that can perform the task of assisting middle school students with their physics lab activities [93]. This smart system uses several advanced technologies, among them the GPT-3.5-turbo large language model (LLM) for natural language processing, YOLOv8 for object detection to find and access the physical objects in the lab, and the latest speech recognition techniques. The first evaluations have shown that the system is on par with experts in the correctness of the facts and even beats them in speed, especially compared to the more powerful GPT-4 model, underlining its efficient and effective tutoring capacities.

This paper is theoretical in nature and compares conventional Intelligent Tutoring Systems [94], including AutoTutor, and the new transformer-based Socratic dialogue systems. The study delves into the engagement of learners, which includes a learner reflecting on their learning experience, a teacher being able to track students' misconceptions, and technology being in line with the pedagogical principles. The paper's insights help to evaluate the efficiency of various tutoring paradigms in the facilitation of understanding and the achievement of educational outcomes. EduChatbot, created by Maheswari and Nagarajan (2024), is a sophisticated hybrid chatbot model that combines logical interpretation of natural language (NLP) and deep learning techniques which allow students to understand their educational queries efficient feature extraction easily, it incorporates XLNet and BERT simultaneously and employs an E1DCNN-LSTM classifier that has been personalized using a novel algorithm called PA-BEPOCPA [95]. As a result of this excellent performance, the bot can surpass the baseline models for the student intent recognition task by up to 62% in accuracy. Therefore, it becomes evident that one of the factors that this technology can leverage to facilitate students' communication with educational systems is the deployment of such models.

2.2.3. Deep Learning-Based Assistants

Deep learning VA utilize deep learning neural networks to analyze and provide answers that are similar to a human. This feature permits them to reach advanced levels of precision and enables user-friendly, situational interactions [96]. In contrast to conventional methods, the deep learning virtual assistants are capable of extracting patterns from large datasets, thereby rendering them versatile and effective in different areas such as education, healthcare, customer service, and intelligent environments.

Chiu et al designed a VA powered by artificial intelligence that is sensitive to emotions, and it uses a DNN architecture. The goal of this assistant is to recognize and understand the emotional states of the users from the different inputs, namely, speech tone, facial cues, and other data, hence delivering the most suitable support [97]. Sophia and Jacob created a student-facing chatbot application utilizing the capabilities of RNNs for text understanding and CNNs for image-based queries or attachments in order to handle the questions [98]. The process description of the flow was also included, indicating that the system initially carries out intent classification via Dialogflow with keyword matching. Next, it employs RNN-based sequence classifiers for semantic intent resolution and, as with CNNs for images, such as photos of homework or diagrams, to provide context-aware responses or simplify the user journey to the correct resources.

Zadeh and Alaeifard designed a hybrid deep learning assistant that integrates 1D CNNs alongside attention mechanisms to identify user sentiment and keep conversational context [99]. To specify more clearly, the attention layer evaluates not only the most recent dialogue history but also user preference embeddings together with sentiment detection. This feature makes it possible for the system to produce replies that can be changed on the fly, both in terms of style and substance. For instance, a user who is upset may receive a comforting message, while a confident one can get brief instructions. The developers of a project designed an AI chatbot that could give suggestions related to shopping, especially for laptops, utilizing an encoder-decoder generative RNN-based model [100]. The assistant manages customer inquiries and picks the correct answers. It generates personalized response messages for users and makes the necessary clicks available for the complete specifications of the laptops. Part of the system, the RNN, was adjusted using product review corpora to make the output more relevant and natural. Burri et al. developed a VA with a strong emphasis on health, aimed at making the patient triage process easier and providing medical information [101]. This assistant uses sequence-to-sequence LSTM networks, which are improved with Word2Vec embeddings. The primary function of the assistant is to provide healthcare recommendations [102], referrals to doctors, and briefings for educational purposes by analyzing the symptoms of users with their history. Moreover, the assistant changes its mode and level of detail depending on the user's educational background and previous interactions.

Apriyanto et al. developed a speech recognition unit for VAs that combines CNNs with RNNs to efficiently understand the audio signal and convert the speech into written language [103]. CNN obtains the spectral features from the spoken words, while the RNN keeps track of the temporal dependencies for the context-based speech recognition. With this combined model, the word error rate is lowered, and a voice control system for a lecture or house learning session can be easily used in real-time [104]. A university project has reengineered the Seq2Seq LSTM chatbot, combined with Word2Vec for intent generation and response generation [105]. The report highlights that the model is capable of handling simple conversational turns (for instance, greetings and FAQs) but confuses semantic ambiguity and cannot handle new query structures. Besides, the model exhibited some weaknesses in keeping the context and frequently gave generic answers in situations where the amount of the training data was not enough, thus making the LSTM-based assistants, with no extensive fine-tuning setup, prone to such problems.

2.2.4. Hybrid Assistants:

One of the latest developments in virtual assistant technologies is the modelling of hybrid systems that provide a better user interaction and more functionalities. Ponmalar et al. developed a VA that combines NLP with the use of BERT for a more efficient text classification and intent detection [106]. Voice Interactive (IVR) interface of the system allows a seamless switching between voice and text communication. The system employs NLP for speech processing and BERT to comprehend the user's intent, thus facilitating easy navigation and smart call-routing by the user, which in turn, can significantly alter the working of an academic help desk.

In [107], research was also conducted on hybrid task-oriented dialog systems that combined fine-tuned BERT models with an intent prediction module. Such a design utilized question answering and classification-based intent recognition, which led to higher accuracy when dealing with synonym variations and yes/no questions in comparison with conventional single-model systems. These changes, taken collectively, indicate the increasing concentration of the user experience and the feasibility of their use in various teaching and service situations.

Another notable example is a prediction assistant that combines LSTNet for deep learning time series analysis with the Prophet model for statistical trend modeling as an initial solution, improved by PSO to fine-tune the hyperparameters [108]. The blending of these techniques gives the assistant deeper forecasting capabilities that be used to anticipate the trends of users' behavior on learning platforms.

2.2.5. Summary Table for VAs

Table 2-2 offers a detailed overview of the related studies. The table illustrates the pros and cons of each cited method, facilitating a straightforward comparison for reference.

Table 2-2: Comparison Between the Chosen Methods for VAs

Approach	Advantages	Disadvantages	Performance Insights	Benefits of T5, EPO, and IoT
Transformer-based Virtual Assistant [76], [77] [78], [79] [80],[81], [82],[83] [84],[85] [86],[87].	- Enhanced Context Understanding significantly surpasses traditional models. - Parallel Processing enables quicker training. - Achieves state-of-the-art results in NLP Tasks. - Utilizes Transfer Learning and Pre-trained Models. - Supports Multi-modal Capabilities.	- Significant Computational and Memory Costs. - Requires a large amount of data. - Pre-trained Models may exhibit Bias.	- Offers exceptional NLU, contextual insight, and fluent responses; nonetheless, it faces challenges with long-term memory retention and latency issues.	- Aims to reduce latency, enhance memory retention, adapt responses in real-time, and lessen bias.
NLP-Based Assistants	- Enhances User Experience: Provides human-like interactions for improved	- Struggles with Context Retention. - Exhibits Bias in Responses.	- Highly effective and adaptable; however, challenges persist	- Focuses on decreasing latency and computational burden, allowing for

[88],[89], [90].[91] [92].[93] [94],[95]	engagement. - Automates Services. - Scalable: Able to support numerous users concurrently. - Involves Continuous Learning: Progresses over time through machine learning (ML) and user feedback loops.	- Associated with High Computational Expenses. - Lacks Common-Sense Reasoning.	with context retention, computational demands, and bias reduction.	dynamic adaptation of the assistant.
Deep Learning-Based Assistants [97].[98] [99].[100] [101].[103] [105].	- Offers Higher Precision and Natural Responses. - Demonstrates contextual understanding. - Scalable alongside Continuous Improvement. - Supports Multilingual Capabilities.	- Carries high Computational Expenses and Latency. - Might embody Bias in Training Data. - Encounters Context Retention Issues in Extended Conversations.	- Exceptionally intelligent, context-aware, and scalable solutions but struggles with computational and contextual constraints.	- Adaptive, reducing computational load and inference latency, while improving context retention and personalization.
T5 (Text-to-Text Transfer Transformer)	- Excels in Language Processing. - Beneficial Pretrained Generalization. - Offers Multilingual Support. - Displays Contextual Awareness.	- Involves high computational costs. - Faces difficulties with long-term memory retention	- Provides superior NLU, yet struggles with resource utilization.	- EPO enhances T5's resource efficiency, resulting in quicker response times.
EPO (Emperor Penguin Optimization)	- Optimized Model Efficiency. - Approaches Faster Training and Inference. - Incorporates Adaptive Learning. - Focuses on Bias Reduction.	- Requires ongoing feedback and real-time adjustments.	/	/
IoT (Internet of Things)	- Enables Real-Time Context Comprehension. - Delivers Personalized Assistance. - Improves Mechanization. - Provides Data-Driven Insights.	- Connectivity issues might influence real-time functionality.	/	/
T5-EPO-	- Features Advanced	- Entails High	- T5 thoroughly	/

IoT virtual assistant (proposed)	Natural Language Understanding (NLU). - Real-Time Optimization through EPO. - Displays Dynamic Adaptability with IoT. - Enhances Response Accuracy. - Promotes Scalability and Customization. - Ensures Efficient Resource Utilization.	Computational Costs. - Faces Challenges with Real-Time Processing. - Presents Training Complexity.	analyzes and formulates responses; EPO reduces computation time, and IoT enhances user focus through immediate adjustments.	

2.3. Works related to Engagement Detection

To create an adaptive and engaging learning environment, it is essential to track, in real time, the involvement of students. Usually, conventional methods rely on the data reported by the students themselves or a simple sensor monitoring that may not be accurate and effective enough. In contrast, the use of computer vision and deep learning has brought about radical changes in the educational sector. One of the most important breakthroughs, specifically in CNNs and the attention mechanisms like Transformers, has allowed the building of more complex systems that can not only engage students but also measure their engagement level.

2.3.1. Computer Vision-Based Approaches

Computer vision techniques utilize multiple visual clues—like facial expressions, eye gaze direction, head pose, and total body posture—to provide detailed assessments of students' focus and engagement. Such methods are very effective in various types of class settings, for instance, in the case of a live online class, or when the video of the class is used for the reading. Researchers have brought in new ways to use hybrid models to get the engagement features from visual data as one of their recent contributions.

Using these cutting-edge techniques, schools can identify and fulfill students' needs instantly. Such a method not only enhances educational service but also facilitates the identification of students in need of assistance or lacking interest, thus providing support. The need for digital media to become completely engaging and to adapt quickly to different learners' needs is the boundless potential that researchers' continuous efforts in perfecting these methods bring about, finally increasing educational achievements.

For instance, Murshed et al. proposed a CNN-driven method to measure kids' interest in digital learning setups [109]. The core of their method lies in the utilization of deep learning models for facial expression recognition, trained on simulated online class data. The CNN performs the task of recognizing indicators of involvement or lack thereof instantly by examining the discussed visual cues, e.g., smiling, frowning, or neutral faces, thus providing a way of non-intrusive supervision by teachers. Zhu et al. employed Vision Transformers (ViTs) to recognize students' feelings by analyzing their facial expressions in teaching environments [110]. Through the implementation of the self-attention mechanism in ViTs, their model was able to surpass standard dynamic facial recognition methods, thus leading to emotion classification with a higher accuracy rate. This type of usage fits perfectly in the intelligent classroom environment, where facial data with high variability can cause problems for CNN performance.

In article [111], the authors presented an energy-efficient Convolutional Neural Network (CNN) design specifically tailored for live engagement detection in remote learning scenarios. The model nearly reached the same precision as more complex CNN models; thus, it was perfect for devices with limited power, such as tablets or school laptops. Their research focus was on the compromise between the processor's capability to perform the task and the reliability of the recognition, which is a critical issue in the application of teaching halls with a large number of students. Tang et al. designed an improved Fast Region-Based CNN (FRCNN) model that integrates the posture features for the improvement of student engagement classification [112]. Their emphasis on low-resolution images not only enhances the flexibility of the model with different video qualities but also makes it suitable for online classes. The inclusion of features such as the position of shoulders and the behavior of leaning provides the classifier with a more in-depth understanding of the involvement that cannot be achieved only by facial cues.

Sharma et al. created a smartphone-driven tool for the assessment of students' involvement through the pictorial record of routine activities, which are accessible via a camera-centric dashboard interface [113]. The application implements the Viola-Jones [114] method for head localization and utilizes a binary classifier to decide the area of the face and the facial features. Besides, they implemented questionnaires to link the visual parameters to the self-reported levels of engagement. Nezami et al. created a dataset of student engagement and implemented the VGG-B CNN model to perform a binary classification of engagement (engaged vs. disengaged) [47]. Their method focused solely on the features of still images and emphasized the need for the availability of reliable datasets for efficient model training. He and Gao decided to merge CNN and LSTM (Long Short-Term Memory) models to capture both the spatial (image-based) and the temporal (behavior over time) aspects of online learning engagement [115]. Their model utilizes student activity logs as well as visual features, enabling a time-aware analysis of learning behaviors. This combination of architectures significantly improves the understanding of students' interactions by correlating facial features with learning behavior patterns.

2.3.2. Physiological Signal-Based Approaches

Engagement detection based on physiological signals is a groundbreaking method that involves using multiple biometric data points to monitor the cognitive and emotional states of learners instantaneously. Some of the leading physiological signals that have been considered include electroencephalography (EEG), heart rate variability (HRV), galvanic skin response (GSR), skin temperature, and respiratory rate. This method seeks to reflect the correct mental challenge, misunderstanding, and engagement more than what can be achieved with a mere visual system by using these internal cues.

A comprehensive review by Bustos-López et al. delved deeply into innovative wearable technologies aimed at detecting the engagement level of learners in educational settings [109]. The survey inspected the most common physiological signals measured, such as pulse, skin conductance, breathing rate, oxygen saturation, electrocardiogram (ECG), and skin temperature changes. Besides physiological signals, the study also discussed a wide range of both commercial and non-commercial wearable devices that use these modalities for sensing and are further effective in detecting different psychological states such as stress, surprise, attention, and anger. Apicella et al. developed a sophisticated and wearable brain-computer interface (BCI) specifically tailored to the real-time monitoring of the EEG signals of students [116]. The primary purpose of this novel apparatus is to measure the cognitive load and affective conditions of the learners, thus providing the possibility for educational interventions to be used right in the actual classroom environment. This not only prolongs students' engagement with the material but also ensures that they receive the most suitable instruction at any given time.

Carroll et al. studied the efficacy of a multi-modal physiological monitoring system that integrates EEG recordings and heart rate variability (HRV) analysis to measure participant engagement in a simulated as well as a live training setting [117]. Such settings encompass the scenarios of first responder training and UAV systems training. Their findings indicated a significant potential for differentiating various engagement stages, thus uncovering the intricate interactions of learners with training material. The seminal research by D'Mello and Kory laid out a crucial real-time model for the identification of the affective state of a user that combined different modalities such as EEG, skin conductance, pupil dilation, and heart rate measurements [118]. The use of multimodal channels in this effect detection system represents an advancement in identifying numerous affective states such as disinterest, anger, misunderstanding, and involvement. Consequently, it opened up the possibility for the design of emotionally and cognitively adaptive learning systems.

In a study, Ananthan et al (2024) used Empatica E4 devices to measure electrodermal activity (EDA) and the inter-beat interval (IBI) from 23 students during actual teaching sessions in the school [119]. It was learned that the students' physical response behaviors remained consistent in varied tasks, and this provides a key signal for tailoring education that is truly capable of changing unique learning profiles. Finally, Singh et al. (2024) presented VisioPhysioENet, a multimodal framework that marries visual cues with physiological signals, particularly cardiovascular features obtained via the plane-orthogonal-to-skin method [120]. Their study showed a notable increase in accuracy of about 8.6% compared to unimodal models on the DAiSEE dataset, thus emphasizing the promise of multi-modal data fusion for engagement detection.

2.3.3. Behavioral and Interaction-Based Approaches

Behavioral and interaction-based methods measure the extent of student engagement by examining various digital activity trails that may include speech features, language input, typing styles, and interaction logs gathered during the execution of learning tasks [121]. Such methods apply significantly to remote and mobile learning systems, as recording the usual video or biometric data is quite challenging. The goal is to understand students' behavior, interaction, and communication with learning content over an extended period.

Fahad et al. proposed a novel approach to identifying emotional engagement that involves the examination of vowel-like regions (VLRs) in the speech signal [122]. These areas, which are the fundamental frequency of the vowel segments, were used to obtain the main acoustic features of the speech, like tone, pitch, prosody, and phonetic rhythm. The researchers, thus, accessing the emotional indicators, were able to ascertain the emotions of the learners in their spoken exchanges. The mentioned research is essential, especially for the areas of speech-activated systems and voice-automated virtual learning environments. In [123], the authors implemented Natural Language Processing (NLP) methods to analyze the students' written input and text-based interactions in a digital learning environment. The approach to analyzing student content involved evaluating the linguistic complexity, semantic richness, and sentiment patterns in a variety of text formats produced by students, including reflective writing, chat messages, and discussion posts. Consequently, the system was able to forecast engagement levels and learning evolution over time.

Bixler and D'Mello examined the capabilities of keystroke dynamics, such as typing speed, rhythm, latency, and pauses, to identify boredom, engagement, and neutrality while writing [124]. They found that changes in typing behavior were tightly associated with cognitive load and affective state; thus, implying that the keyboard can become a sensor for learner emotions. Bosch et al. introduced a multimodal behavioral model that combined log data (e.g., clicks and time-on-task) with linguistic features derived from forum posts for engagement prediction [125]. Their findings revealed that the use of both digital behavior traces and written discourse led to the improvement of the model's accuracy, thus indicating that multisource behavioral data can facilitate student modeling.

2.3.4. Summary table for Engagement Detection

Table 2-3 offers a thorough overview of the relevant literature. It highlights the strengths and weaknesses of each method discussed, providing an accessible comparison for reference.

Table 2-3: Comparison Between the Chosen Methods for Engagement Detection

Category	Advantages	Disadvantages	Contribution to Engagement Detection	Limitations	Why CNN-ViT is Better
Computer Vision-Based Approaches [109], [110], [111], [112], [113], [114],	- Non-intrusive and relies solely on a camera. - Implemented in real-time. - Effectively assesses engagement through facial expressions and eye contact.	- Performance is dependent on lighting conditions and camera quality. - Requires extensive datasets for training.	- Gauges engagement through facial characteristics, head positioning, and gaze. - Performs effectively in video-based learning environments.	- Limited in assessing deep cognitive involvement. - Restricted to interactions based on facial features.	- CNN-ViT addresses the limitations of CNNs by incorporating global context awareness through ViT, improving feature representation.

[47], [115]					
Physiological Signal-Based Approaches [109], [116], [117], [118], [119], [120].	- Offers authentic physiological indicators of engagement. - Highly precise across various channels.	- These methods require specialized equipment such as EEG headsets and smartphones. - Privacy considerations are still significant.	- Monitors immediate engagement through biological signals. - Valuable for conducting accurate assessments of both cognitive and emotional states.	- Ineffective in crowded environments, suitable only for one-on-one or small group interactions. - Not suitable for online collaborative reflection (OCR) without the necessary electronic devices.	- A key advantage of CNN-ViT is that it is non-invasive and only needs video processing, making it applicable for broader educational contexts in virtual environments.
Behavioral and Interaction-Based Approaches [122], [123], [124], [125].	- Non-intrusive and easy to set up using digital platforms. - Functions effectively in online learning frameworks. - Capable of tracking engagement patterns over time.	- Limited to recognizing engagement during interactive activities. - Cannot assess facial expressions or emotional engagement signals. - Exhibits high variability among users.	- Detects engagement through typing, scrolling, and mouse movement patterns. - Useful in adaptive learning strategies.	- Unable to determine passive engagement in video-based learning. - Fails to capture emotional or cognitive states.	- CNN-ViT analyzes both passive and active engagement through facial expressions and gaze patterns, making it suitable for a variety of learning contexts.
Proposed Hybrid CNN-ViT Model	- Combines spatial feature extraction (CNN) with global attention modeling (ViT). - Non-intrusive and scalable. - Operates in real-time with high accuracy. - Captures both local (micro-	- Requires large labeled datasets for training. - Computationally intensive.	- Extracts detailed engagement features using CNN. - Understands long-range dependencies with ViT. - Ideal for video-based learning in online collaborative		

	expressions) and global (context-aware) engagement patterns.		reflection (OCR).		

2.4. Works related to Progress Prediction

Predicting student performance is the basis of one of the core personalized learning features that aims to design educational paths suited to the needs of every learner. Usually, these predictions rely on classical statistical models and decision trees. Although these methods have succeeded in several cases, they still often fail to reflect the complex nature of students' learning processes, which can be affected by numerous factors such as family background, individual learning preferences, and motivation.

Recently, the emergence of more sophisticated techniques has been notable, especially those related to machine learning. Among those, RNNs and their types, such as GRUs and LSTMs, have attained significance. The main advantage of these models is their ability to handle sequential data with great efficiency, and thus, they are very suitable for all the temporal facets of learning, for instance, the student's interaction with the content over time. A study suggests that RNNs can very finely capture the student learning trajectories, which is key to understanding how performance changes with different instructional methods.

2.4.1. Statistical Approaches for Student Progress Modeling

Traditional statistical methods have had an influence on the field of education research. Methods such as regression analysis, analysis of variance (ANOVA), and t-tests are used frequently to determine the connections between different learning behaviors and educational outcomes. Moreover, these methods enable the discovery of significant trends and relationships, thus enabling researchers to pinpoint and measure the impact of educational interventions. An example, regression analysis may indicate the association between educational strategies and student performance metrics, while ANOVA may be applied to assess the efficiency of different pedagogical approaches over various populations.

These traditional models offer insights of true value that influence policy decisions, lead the formation of efficient instructional strategies, and recognize possible risk factors in learning environments, which constitutes a considerable contribution to their understanding of educational processes and the outcomes thereof. Moreover, they can combine the qualities of both traditional statistical methods and modern machine learning frameworks to achieve the best results in predicting students' performance. As a result, educators and researchers will be able to provide personalized learning experiences that cater to the diverse needs of individual learners.

For instance, the authors used a Bayesian hierarchical growth curve model to explore the academic development over time of 770 students [126]. This technique enabled them to represent the variances not only at the individual level but also at the module level, thus unearthing some hidden factors that lead to academic progress and are hardly ever revealed by the typical statistical models. The technique also allowed for the incorporation of nested structures, which are prevalent in education settings such as students within classes and classes within schools, by using hierarchical priors. Similarly, McNamara et al. utilized a Bayesian inference setting to figure out which features derived from traditional and AI-based predictors would better forecast the academic performance of STEM undergraduates [127]. They essentially combined preliminary assumptions about students' success with actual data, thus obtaining practical suggestions about students' staying and leaving the institution. The authors also emphasized that the adaptability of the Bayesian system was very advantageous for dealing with doubt scenarios and changing one's opinion with new data.

Despite their advantages, conventional statistical models might not be able to efficiently handle high-dimensional behavioral data or situations where instant forecasts are required. Thus, to increase the usability and the efficiency of the models in the setting of targeted interventions, they are often supplemented with optimization and machine learning. Through the use of gradient boosting optimized logistic regression, Maier & Klotz developed an adaptive learning system to identify students who require individualized feedback. Their method represents a blend of the transparency of traditional techniques with the capabilities of cutting-edge AI [128]. Sun used Hierarchical Bayesian Knowledge Tracing with a group of undergrad engineering students in a Statics class [129]. The model allowed Sun to see not only the differences between students but also the variations at the topic level, uncovering the skills progression that learners had and differentiating the learner groups, hence, facilitating the tailored learning that the instructor had.

The Bayesian hierarchical model was also chosen to handle the data describing the educational achievement of 630 South African STEM undergraduates over five years, from 2019 to 2023 [130]. The STEM study implemented the Bayesian hierarchical model to perform a comparative analysis of different programs employing partial pooling to adjust for variations between different academic programs and years. The study revealed that mathematical skills and economic status were two vital predictors of academic outcomes, thus pointing to the issue of fairness in education. Another study had implemented Bayesian hierarchical modeling as a method to investigate the relationship between matriculation results in various subjects, including mathematics, psychology, and health education, and university performance in the following ten years [131]. The research not only measured the influence of the exam results on future educational achievements but also depicted nested data structures such as the faculty, year, and student.

2.4.2. Machine Learning-Based Methods

Machine learning techniques examine both current and past student information to identify trends by which students' strengths and weaknesses are identified, along with the learning areas that require targeted improvement [13]. These methods achieve higher and higher accuracy of predictions by regularly changing the model weights when new data comes in. For example, peer-assessment data has been employed to train the models that are able to predict student performance. Afterwards, these models were compared with the ones established using demographic data, grades, and test scores. The models based on peer assessment signaled the opportunity of using collaborative learning artifacts for the more accurate prediction of academic outcomes, thus being more sensitive to students' real-time engagement [132].

One of the ensembles learning methods, the XGBoost algorithm, has also been applied to the discovery of the retention, succession, and graduation patterns datasets [133]. To trace the relations between the early-stage features and the expected student outcomes, XGBoost builds on the principle of combining several weak decision tree learners. Its effectiveness and precision make it a perfect tool for handling big and varied educational datasets. The research team pointed out that this method could be useful in identifying at-risk students at an early stage, so the planned interventions in the open classroom and blended learning settings could be administered on time. The integration of Virtual Learning Environment (VLE) data with student record datasets in a Management Information System has enabled the use of Random Forest (RF) classifiers for the prediction of student progression [134]. RF models are useful, especially for feature selection, because they can list a large number of features according to their importance. This function makes them very useful in an educational setting where there are both behavioral and performance metrics. By studying VLE activity combined with grades and demographic factors, these models cannot only find the easiest way to lead the students to academic success but also can discover the most hidden indicators of the same.

Evolutionary computation methods like genetic programming have been used to develop prognostic models that rely on the learning and evaluation data of students [135], [136]. One of the characteristics of genetic programming is its ability to figure out the complex relationship between different features, such as class participation, knowledge acquisition, and summative performance, through successive generations of model structure evolution. The method is also capable of cross-validating its results by comparing predictions with independent datasets from various online courses, which has proven its high potential for generalization. Wang and Yu developed a model that relies on the Behavior Indicator-Driven Logistic Regression innovative approach [137]. This method represents a logistic regression model trained with behavioral indicators selected by the study to predict the performance of a student. In this work, the authors achieved this by extracting eleven key metrics of online learning behavior and employing correlation-based feature selection. Hence, this model was able to substantially eclipse baseline models, thereby indicating that the timing and the proactive nature of the learning behavior can be valuable for prediction.

2.4.3. Deep Learning-Based Methods

Deep learning, a subset of machine learning, utilizes multilayered neural networks that are capable of grasping complex and nonlinear data representations. These networks can combine numerous inputs, including attendance records, behavior metrics, and test scores, to identify hidden relationships that impact students' progress. It is their feature to handle unprocessed, high-dimensional, and multimodal data that makes them ideal for adaptive learning environments.

GritNet, as described in [138], is a deep learning model that leverages Bidirectional Long Short-Term Memory (BLSTM) networks. The model used to predict student progress as a sequential event forecasting problem, where the sequence and time of the student's learning activities matter a lot. GritNet, through the chronologically ordered student activity logs, can grasp the temporal dependencies that occur both before and after the student's learning process. This combination of directions not only extends the representation of the student's gradual knowledge acquisition but also facilitates the recognition of disengagement symptoms at an early stage. Moreover, this type of sequential modeling becomes the basis for the accurate prediction of academic outcomes in courses. In [139], scientists applied a Back

Propagation Neural Network (BP-NN) to forecast academic performance that can divide students into classes. BP-NNs are supervised learning models that, through error backpropagation, continually adjust weights to minimize prediction errors. The model is trained using student scores, involvement rates, and behavior metrics, and thus generates classification outputs that are indicative of the levels of performance. This method, though pertinent, is considered less advanced compared to that of recurrent or attention-based architectures since it lacks the feature of explicit time dependencies modeling. Nevertheless, it conveys the idea that even basic feedforward architectures might wield significant effectiveness if synergized with high-quality educational datasets.

The research in [140] utilized a fully connected ANN as a tool to anticipate the results of a student in an engineering course. Some of the inputs for the model were the student's attendance rate, assignment grade, and academic history. The neural network was trained to discern non-linear relationships among the given features, thus offering a probabilistic evaluation of achieving a successful performance in the course. Although this strategy is more concentrated on static predictors as opposed to time-varying behavioral data, the achievement of ANN-based methods in the identification of risk and the provision of targeted academic intervention remains. Peng et al. developed a Graph Transformer Architecture (GTA) aimed at employee performance forecasting in the collaborative learning network [141]. GTA fuses transformer-based self-attention features with graph-based relational modeling, thereby enabling the network to check both individual behaviors and social interactions. This combination method is extremely effective in capturing the social learning dynamics, thus making it very suitable for group-based online courses or project-based classrooms. The model has the most similarities with deep learning-based student performance prediction, particularly in collaborative settings.

In [142], the authors used a Gated Recurrent Unit (GRU) network to make student performance predictions as early as possible. GRUs, being a type of LSTM, are designed to use less time and storage than regular LSTMs while still capturing the temporal characteristics of the data. The researchers combined different kinds of sequential data, e.g., engagement metrics, assessment results, and attendance logs, to identify students in-danger of dropping out before the end of the course. This method is very close to the practice of early-warning systems in educational data mining, which is their direct application. Delianidi et al. presented a dynamic neural network framework that can represent changes in learning behaviors throughout time. The main difference with static prediction models is that this method adjusts to changing user patterns by constantly integrating new data [143]. Moreover, the network exploited the educational program, grading landmarks, and participation to infer the endpoints and recognize the students who are to be helped. Such a technique can be very useful for long courses or those with a variable pace of study.

2.4.4. Data Mining & Educational Analytics Approaches

Data mining and educational analytics are two areas that essentially deal with finding useful inputs from large and complex educational datasets. However, through the application of various analytical techniques like clustering, predictive modeling, and association rule mining, educators are now able to recognize student behavior motives; they can also forecast the academic progression of students and thereby initiate a prompt intervention. If used appropriately in the teaching and learning settings, these tools can serve as a guarantee for improved student participation, continuation of studies, and achievement of results.

Willey and Gardner examined the use of SPARKPLUS, an online platform designed to enhance peer and self-assessment of group work [144]. SPARKPLUS allows students to grade their peers' contributions, creating quantitative weightings that enable fair adjustments to the overall group grades. Beyond serving as a grading tool, the platform also tracks detailed metrics on participation and collaboration, helping teachers identify issues such as uneven workload distribution or disengaged group members. By utilizing the data generated by learners, teachers can better understand the group's dynamics. Additionally, it encourages accountability, reflection, and skill development in effective teamwork. Another method used the psychological contract model as a tool to foresee the probability of students leaving [145]. The researchers include students' initial educational hopes and compare them with their actual academic experiences, identifying expectation-reality gaps which, in their view, could cause disengagement of students. The data were gathered through structured interviews with underachieving students, and to determine the causes, patterns were analyzed, which might include lack of support, poor social integration, or curriculum misalignment. Due to the identification of risk factors and the subsequent targeted interventions, not only academic performance but also retention has increased. This method demonstrates how combining qualitative insights with predictive analytics can uncover the complete profile of students' risk factors.

Romero et al. have combined clustering techniques with VLE log data to discover student groups whose members have similar behavioral patterns [146]. Those clusters, for example, high-frequency users, deadline rushers, and minimal engagers, were linked to final grades by teachers to predict which groups would have difficulties. In this instance, the prediction of struggling students is achieved through data-driven analysis, rather than by using students' self-assessment information. Besides, this method is scalable for large cohorts as it does not require direct student input. Jayaprakash et al. have designed a groundbreaking early warning predictive analytics dashboard that employs logistic regression and decision trees for data processing from Learning Management Systems (LMS) [147]. Besides demographic information, the data comprises login frequency, assessment scores, and discussion participation. The dashboard is designed to deliver instant notifications to faculty members, indicating those learners who are highly probable of failure or getting disengaged. Consequently, the system empowers timely intervention methods, supported by data, and corresponds closely to approaches such as the psychological contract framework, which, through automation of risk detection at the institutional level, facilitates the seamless implementation of risk.

2.4.5. Hybrid Models

Hybrid models combine traditional statistical methods with machine learning and deep learning that leverage the advantages of each method, resulting in enhanced and more practical predictions in the field of education. This power duo demonstrates the capability of handling any data, be it structured demographic information or unstructured behavioral logs. Besides, it can also detect the temporal dependencies and non-linear relationships within the students' performance data. Such a combined method is very efficient for the large educational analytics market, as it can use both historical and present data, thus enabling the anticipatory interventions and personalized learning strategies.

Kukkar et al. created a combined forecasting system that integrates RNNs with LSTM structures, supported by machine learning methods for feature engineering and tuning [148]. The researchers had academic records, VLE interactions, and participation metrics in their dataset and thus were able to exploit the RNN-LSTM’s capacity for holding temporal dependencies in student activity. Besides improving the model’s interpretability and predictive performance with feature selection methods, a significant contribution to the accuracy of longitudinal progression has been made.

Ouyang et al. developed a clever, AI-driven approach to improve the educational outcomes of students in an online engineering course [149]. They performed a quasiexperiment, where a single student group used AI-based predictive models to provide feedback. These models combined neural network architecture with traditional performance indices. The joint feedback system examined areas such as participation, cooperation, and past performance, and enabled the user to receive particular tips on how to improve their performance. The findings showed that there was a statistically significant increase in the students of the AI-assisted group in terms of satisfaction, collaborative learning behavior, and total course achievement as compared to the control group. Waheed et al. brought forward a combined model that merges Random Forest and DNN (RF-DNN) for the forecast of the students who may fall into the risk category in higher education [150]. Random Forest was utilized to clean the data and prioritize features by their significance, thus efficiently reducing noise and dimensionality. After that, the most significant attributes were chosen for the deep neural network to carry out the definitive forecasts. Such a multi-step method essentially combined the interpretable part of tree-based models with the efficient feature extraction part of deep learning, thus allowing high recall rates for those students who required an early intervention to be captured.

2.4.6. Summary Table for Progress Prediction

Table 2-4 offers a detailed overview of the efforts related to predicting student progress. It lists and contrasts the pros and cons of each approach, emphasizing the improvements made in comparison to our model.

Table 2-4: Comparison Between the Chosen Methods for Progress Prediction

Category	General Advantages	General Disadvantages	Performance Insights	Performance Limitations	How HRT-GRU Improves
Traditional Statistical Techniques [126], [127], [128], [129], [130], [131]	- Simple and easily comprehensible - Requires minimal computational resources	- Presumes linear relationships - Struggles with processing high-dimensional information	- Clear decision-making process - Works efficiently with small datasets	- Inadequate for capturing the complexities of learner behaviors - Not suitable for enhancing step-by-step or disorganized data	- Facilitates deeper analysis through hierarchical reasoning - Integrates the GRU model with time-oriented learning

Machine Learning -Based Methods [132],[133], [134], [135], [137]	- More effective with non-linear data - Able to identify complex relationships	- Requires feature engineering - Some models have issues with interpretability	- Compatible with organized student data - Offers significantly better accuracy compared to statistical models	- Has difficulty with sequential patterns - Requires large labeled datasets for training	- Utilizes transformer-like hierarchical reasoning for better feature extraction
Deep Learning -Based Methods [138], [139], [140], [141], [142], [143]	- Ideal for large-scale unstructured datasets - Automatically understands temporal dependencies	- Requires a large dataset - High computational costs	- Understands long-term patterns in student learning trends - Effective with sequential student data	- Challenging to analyze outcomes - Can overfit with a small amount of data	- Uses GRU for sequential learning while HRT enhances interpretability
Data Mining & Educational Analytics Approaches [144], [145] [146], [147]	- Extracts significant patterns from student data - Capable of working with structured, semi-structured, and unstructured data	- Lacks predictive power - Requires extensive preprocessing and expert insight	- Helps identify at-risk students early - Supports adaptive learning and personalized recommendations	- Often more descriptive than predictive - May overlook sequential dependencies in student learning behaviors	- Merges predictive deep learning (GRU) with hierarchical analytics (HRT) - Moves beyond descriptive analysis to provide actionable predictions
GRU-Based Model	- Manages sequential dependencies effectively - More efficient memory usage than LSTMs	- Lacks hierarchical reasoning capabilities - Less effective with structured/tabular data	- Strong for time-series student data - Recognizes student interaction patterns over time	- Has difficulty with non-sequential dependencies - Limited in handling static features	- HRT enhances reasoning for structured data - Improves interpretability beyond GRU's opaque nature
HRT-GRU Model (Our Approac	- Combination of deep learning and structured reasoning - Recognizes	- More complex architecture - Requires greater computational	- Merges temporal and hierarchical elements for better predictions	- Requires adjustments to balance interpretability with accuracy	

h)	both sequential and hierarchical patterns	resources	of student progress - Adapts well to both individual and collaborative learning	- Involves higher computational demands	

3. Conclusion

This chapter examined the literature contributions to a diverse ecosystem connected with educational technology, particularly OCRs. This variety encompasses several fields, including the implementation of search engines for the retrieval of academic articles, setting up smart virtual assistants to guide the learner, systems that can detect student engagement, and predictive models that assist academic progress monitoring.

Techniques in search engines and data recovery have substantially evolved from the initial simplistic keyword-based matching to the latest semantic and context-aware algorithms. These improvements enable users to access educational resources easily and quickly, thus saving valuable time. VAs, equipped with conversational AI and natural language understanding, have gained traction as interactive learning companions that deliver customized guidance, continue posing inquiries, and encourage the practice of self-directed learning.

Engagement detection research progressed over time, and the use of different data and AI-based models replaced the previous simple activity tracking. Fundamentally, these models rely on a combination of behavioral, interaction-based, and physiological signals to give insights into one's motivation and participation. At the same time, progress prediction research, which relies on machine learning, deep learning, data mining, and hybrid models, has predicted student performance and thus identified the at-risk learners early enough to implement intervention strategies.

These areas of knowledge, taken together, form a body of knowledge that constitutes the technological and methodological base for our work. They exhibit a need for custom, data-based, and situation-aware learning systems that not only comprehend the needs of students but also provide the required help without delay. These various methods were combined to leverage their respective strengths to cover the current shortcomings in a single framework of retrieval, assistance, engagement monitoring, and progress prediction.

Next, we will proceed to the Contribution chapter, where we will outline our innovative, new solutions. This section will focus on the features of the architectural design and the methodological advancements that made it different from other solutions. We will demonstrate how our system changes and, to a great extent, improves the scenario of intelligent and adaptive educational support, thus providing users with a more personalized and efficient learning experience.

Chapter 3: Contributions

1. Introduction

The previous chapter offered a general overview of the current status of intelligent educational systems by presenting methods and models that have been developed to improve online learning environments. In addition, it revealed the research gaps—the drawbacks and the misses of the previous studies, especially related to adaptability, engagement tracking, real-time support, and resource retrieval in Open Classroom settings.

Aiming to discover the best ways to counter the major challenges that OCR environments bring about—issues such as students' lack of interest, limited availability of relevant educational materials, and the requirement of immediate help—we present a set of intelligent, AI-driven solutions that are extremely robust indeed. The aforementioned set of solutions is designed to bring a new level to the educational process by focusing more on adjustability, interaction, and achieved learning outcomes.

We prioritize OCR hurdles head-on by deploying intelligent, AI-based solutions, from search engines to tracking students' progress to virtual assistants. These solutions are designed to empower teachers with tools that can be used online to turn physical classroom limitations into digital opportunities for efficient and effective learning. Each of the solutions we will introduce and explain will be connected to specific scenarios regarding the challenges OCR can encounter, and they will move the field of education to the next level.

2. *Contribution 1: Advanced Text Prediction System Integrated Within the Search Engine for The Open Classroom Approach Based on Particle Swarm Optimization and Long Short-Term Memory Models [151].*

2.1. Description

To mitigate some of the issues of distraction and content overload that OCRs can regularly create, we proposed a sophisticated real-time text prediction system designed to increase the platform's intelligent search capabilities within OCRs. By offering real-time, context-based suggestions that the user can accept, reject, or modify as they type, our prediction system provides evident added value, even just its display of suggestions in educational contexts. So, a user begins to type the letter "A", the prediction system may provide "Apple" as a suggestion. Then, as they type "AR", the prediction system will narrow the suggestions to "art" and "artistic". Then, once the user types "ART", the prediction system may provide suggestions such as "artist" and "artificial". Finally, pressing "space" after the word "artificial", the prediction system will indicate the most likely next word considering the user typed "artificial", and may suggest either "intelligence" or "neural network". Since our text prediction suggestion process appears to take them away from being distracted to navigate through their OCR platform more reliably, with less typing effort, speed increases, and impromptu but relevant recommendations of terms that users may not have been conscious enough to articulate during their engagement in an educational context.

To support this capability, we submitted a unique hybrid model that merges LSTM neural networks and PSO—we term this the PSO-LSTM approach. LSTM neural networks are a type of RNN that performs ideally with sequential data. LSTMs handle long-term dependency issues due to containing memory cells and gates. The memory gate regulates whether to save, update, or forget inputs, and when to do so, benefiting the prediction task. LSTMs are very useful for next-word predictions, in which context and temporal dependencies are critical.

On the other hand, PSO is a strong bio-inspired optimization algorithm that mimics the group behaviors of populations or swarms (e.g., fish and birds) and provides a means to search high-dimensional spaces to discover optimal solutions. In our implementation, PSO allows us to automatically fine-tune the LSTM's critical hyperparameters (number of hidden units, dropout rates, learning rate, etc.) to ensure they accurately and efficiently produce a prediction model without the need for manual trial-and-error fine-tuning. In essence, this hybridization resolves the key weaknesses of deep learning systems, which are their inherent sensitivity to hyperparameter choices.

In this study, we propose an intelligent search engine along with a sophisticated, context-aware text prediction system, designed in real-time and customized for the OCR learning environment. The new system illustrated in Figure 3-1 utilizes a hybrid architecture where LSTM networks handle the complex sequential relationships of user requests, and PSO is used for efficient parameter tuning. The main goal of this architecture is to deliver search results promptly, accurately, and adaptively — not only anticipating user intent but also enhancing the learning process and increasing learner engagement on digital learning platforms.

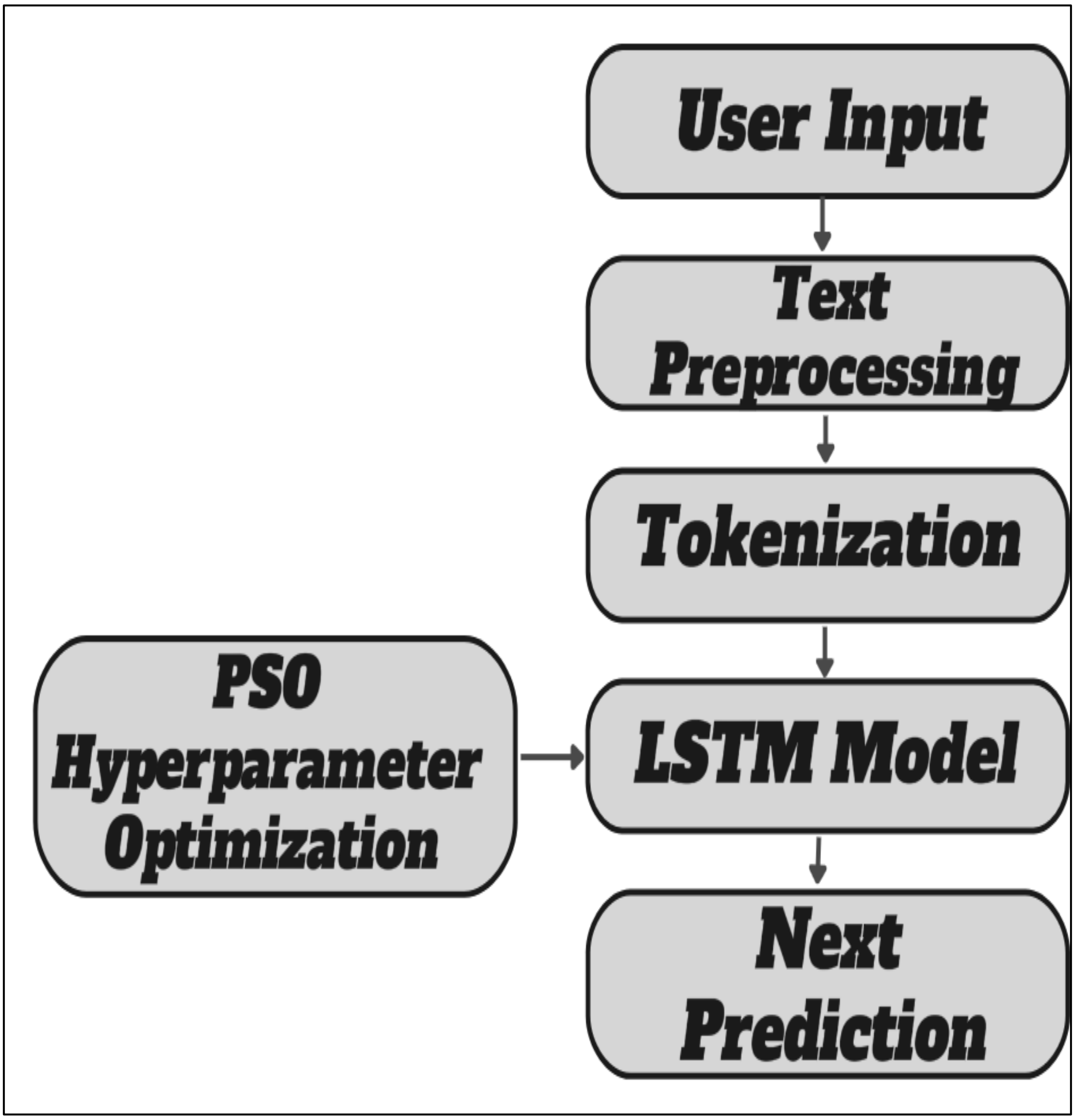

Figure 3-1: PSO-LSTM Workflow Diagram

- Once a user provides an input, the system preprocesses the input string in such a way that it is transformed into lowercase, tokenized, and the relevant characters or tokens are kept while cleaning.
- This token sequence that has been cleaned is next transferred to the LSTM network, which employs it to predict the most likely continuation, which is grounded in the knowledge that the network has accumulated over time.
- Each keystroke serves to update the real-time predictions and also the suggestions list, whose ranks are determined in relation to the user's input changes.
- Whenever the user either presses the space key or chooses a suggestion from the list, the sequence changes accordingly. Thus, the LSTM network will start from a new context when it comes to prediction.

The prediction engine is created to serve the double purpose of character-level and word-level modeling. This paves the way for higher flexibility and support for different writing styles, ranging from casual to formal queries. Moreover, the PSO module is running all the time in the background. It searches for the best parameters and periodically re-trains the LSTM while the new query log files are collected. This procedure ensures that the system is capable of not only learning but also evolving to reflect future language usage trends, and it can work well with each student in the Open Classroom community.

2.2. Methodology

This section describes the process of designing, implementing, and evaluating our intelligent text prediction system for Open Classroom Resources (OCR). Our intelligent text prediction system uses a hybrid model that combines the LSTM networks and the PSO model to provide efficient and accurate word prediction.

We first introduce the architecture of the system, detailing the data collection, cleaning, and loading process for making our data ready for training. We cover the experimental structure and describe the dataset used, the development environment and setup, the training process, and the evaluation. Next, we describe the LSTM model as a sequence-learning measure, and subsequently cover how PSO optimizes the hyperparameters of the LSTM, thereby capturing the temporal dependencies of the data, with minimal manual tuning. By the end, we have established the implementation pipeline of our predictive engine, which will allow us to then analyze the performance of our intelligent predictive text engine in the next chapter.

2.2.1. Development Environment

The LSTM-PSO hybrid model implementation was conducted in a stable development environment having a well-organized dataset to ensure consistent, scalable, and repeatable experiments. The next segment offers the chosen technology stack, development tools, and the data preparation pipeline.

- ***Programming Language:***

Python 3.10 was selected as the primary programming language due to its wide range of modules and support for machine learning libraries. The flexibility of the language is the central focus here, and in addition to that comes the easy integration with the latest data processing frameworks and specific deep learning tools, which promise to provide a smooth workflow for efficient development and, eventually, help users overcome their challenges [152].

- ***Deep Learning Libraries:***

- TensorFlow 2.x with Keras API: The deep learning model was programmed, executed, and assessed through TensorFlow 2.x. Here, Keras was used as an API to achieve a more user-friendly interface that would enable easy and stackable customization of models [153]. This environment is ideal for neural network architectures and is equipped with optimizations capable of accelerating performance and conveniently allowing for on-the-fly adjustments.

- PySwarms: This library has been employed to work out the PSO algorithm, which is the technique that has been best for fitting the hyperparameters of the LSTM model [154]. As the coding structure is the PSO algorithm, the program PSO has an efficient implementation, highlighting parallel processing and rapid convergence to the optimal parameters.

- ***NLP and Data Cleaning Tools:***

- NLTK (Natural Language Toolkit): NLTK is regarded as the basic natural language processing tool that can perform many tasks of tokenization, stopword elimination, and lemmatization [155]. If the data is preprocessed correctly by this tool, then the model results will be more accurate.

- WordNinja: This one is a text processing tool used for splitting words programmatically, especially when dealing with long or complicated terms in the text data. As a result, text representation will have a more detailed or fine-grained meaning.

- ***Data Handling and Processing:***

- Pandas: This effective data-shaping library was exploited to process text data. The library can handle big-sized structured and unstructured datasets smoothly and effectively, and also extract metadata, clean, and reshape it efficiently [156]. With the use of its DataFrame structure, the primary concern of diverse data types was drastically relieved.

- NumPy: Focused on carrying out numerical computations in an optimal way, NumPy has been a major contributor to solving computational concerns [157]. Comparative methods and operations are often used in the step where we are essentially dealing with input data and model training.

- ***Visualization:***

- Matplotlib: We successfully created detailed plots of training metrics, loss curves, and performance comparison charts [158]. The visualization allowed continuous monitoring of the model's performance, enabling us to make necessary adjustments during the experiments.

The establishment of such an intensive and complete development platform, when taken in combination with the picked tools and libraries, actually became the root of the entire LSTM-PSO hybrid model's implementation, thereby allowing a deep and organized way to the machine learning experimentation.

2.2.1.1. Dataset

We used the Arxiv article dataset [159], for our research. This excellent dataset contains metadata of more than 176,000 scientific articles saved on the Arxiv platform. It contains all the necessary information, the list of all articles, subject classifications, publication date, and issues, abstracts, and titles, which are indispensable for conducting thorough analyses of scientific discourse.

Our preprocessing strategy is making the collected data more accurate and effective in the direction of quality by properly separating the dataset for training and testing, a very important aspect in the correctness of our analyses. We begin with text data preparation on a large scale, where we lose non-informative details like URLs, email addresses, HTML tags, special characters, digits, and underscores to control noise. Then, by the use of Natural

Language Toolkit (NLTK), we encode the words to transform the cleaned text into a structured format. Here we go with the separation of concatenated words, also known as Wordninja, the removal of stop words, and the standardized remaining words. We used the Keras Tokenizer to prepare the text for machine learning, where we transform preprocessed text into an integer sequence, each of which corresponds to a unique word; thus, we can better extract the model from the text. Then we set the length of all these sequences to a maximum that is the same for both sets of data, have their sizes adequate for model integration, and hence improve the accuracy of the trained model.

2.2.1.2. Preprocessing Pipeline

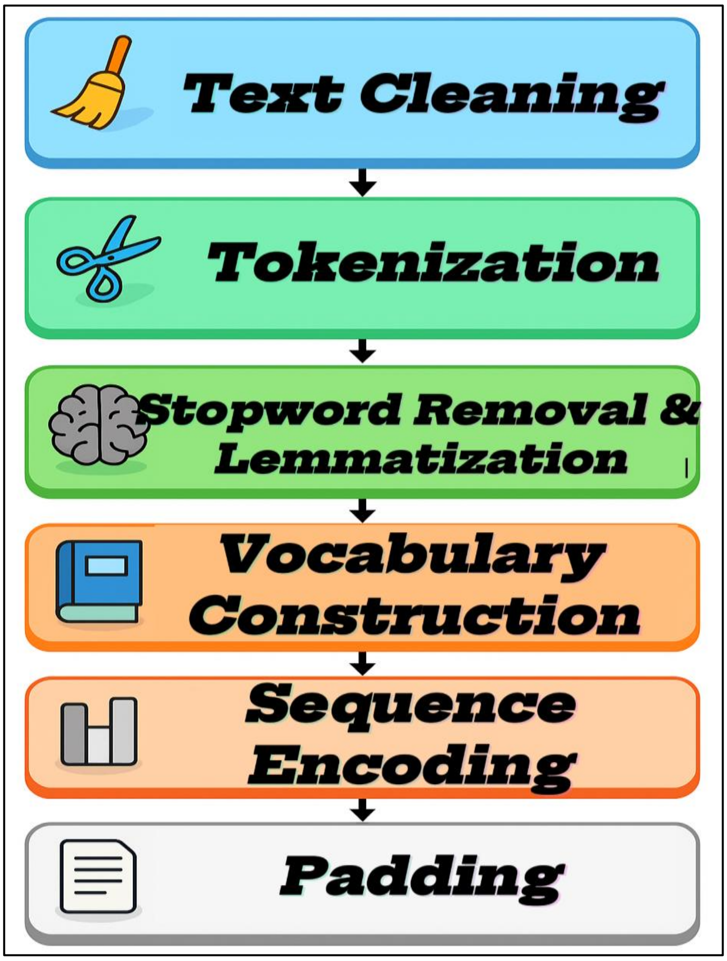

Figure 3-2: Preprocessing Pipeline Flowchart

- ***Text Cleaning:*** Initially, a full-text cleaning process was performed to confirm that the data was of high quality. The process to eliminate noise from the dataset was carried out with special attention to detail. No LaTeX characters, special symbols, numerical digits, hyperlinks, e-mails, or any non-ASCII symbols were left out. To ensure a common normalization standard, all characters were converted to lowercase after noiseless input was obtained. The last step began with the replacement of excess spaces between the words and tokens that illustrated the necessary data aimed for forthcoming processing.

- ***Tokenization:*** NLTK's `word_tokenize()` method was used to accurately and correctly tokenize the words and punctuation marks. Furthermore, WordNinja was used to split compound terms and **camelCase** entries further. By employing this method, it was possible not only to identify words consisting of more than one token but also to cover the case when concatenated words were realized as individual tokens and did not lose their contextual meaning.
- ***Stopword Removal & Lemmatization:*** To increase data relevance, frequently occurring English words were systematically removed with NLTK's extensive stopword list. This trimming of less important words would be an effective technique to give the model a more meaningful role to play in the text. After that, lemmatization methods were used to bring words back to their base forms, for example, the word "running" became "run" as the lemma made vocabulary understanding more generic.
- ***Vocabulary Construction:*** Keras's Tokenizer was used to build the vocabulary for developing a word-to-index mapping most exhaustively. For the simplification of the model, a vocabulary of 10,000 words was opted for, based on the most frequent ones. This has been done judiciously to ensure that the model is still capable of covering a wide range of strong words while being compact, with a smaller vocabulary size facilitating faster computational speed.
- ***Sequence Encoding & Padding:*** Each word is converted into an indexed vector, later used to create a numerical representation for the machine learning process. The length of the sequence is also standard across the input data by using the zero-padding method. That is when sequences were adjusted to a length of 30 tokens, being the maximum limit, and according to the experiment, it proved to be the best time.
- ***Train/Validation/Test Split:*** The dataset was partitioned into three sets: training, validation, and test sets. In detail, the 70% split was used for the training set, and the 30% data split was 15%-15% for validation and testing. By choosing this method, each category was ensured to still have the same distribution of samples,

hence providing a more reliable model performance assessment through this stratified sampling approach.

The highly detailed and exhaustive preprocessing pipeline not only ensured that the model was trained on top-quality, domain-relevant, and linguistically diverse data, but it also provided the LSTM-PSO system with a firm base. Therefore, the system was designed to operate efficiently and swiftly in real-time educational situations, thereby demonstrating its capabilities through highly accurate predictive text generation and becoming an integral part of the intelligent search engine.

2.2.2. Particle Swarm Optimization

Particle Swarm Optimization is a sophisticated population-based metaheuristic algorithm that draws inspiration from the collective social behaviors of natural systems, such as birds flocking and fish schooling. Due to the exceptional efficacy of this algorithm, it has handled various continuous optimization problems, so it has applications in neural network training and hyperparameter tuning with demonstrated success. Here, PSO serves as a strategic means of modifying the LSTM model's hyperparameters to improve the latter's real-time text prediction performance.

In PSO implementation, a parcel of particles is employed to represent selections of potential solutions in the landscape of optimization. Every particle is confined to the given search area and updates its position concerning the following aspects:

- ***Personal Best Position (pBest):*** Each particle keeps track of the best-known position that it has reached in the search space, i.e., the position that has given the highest fitness value till then.
- ***Global Best Position (gBest):*** This is the position that all the particles consider as the best and is the one that gives a measure of the exploration of the search space.
- ***Velocity Vector:*** There is a vector representing velocity, and it tells about the speed and the direction of the particle moving through the search space. It is a vector whose direction and magnitude are dynamically updated to improve the search process.

A single particle's motion is based on a combination of its own experiences and the information it collects as a group member. This dual mechanism gives a push toward an optimality that is possible by the balance of independent self-discovery and assimilation as the two conflicting influences. Figure 3-3 illustrates how the PSO operates.

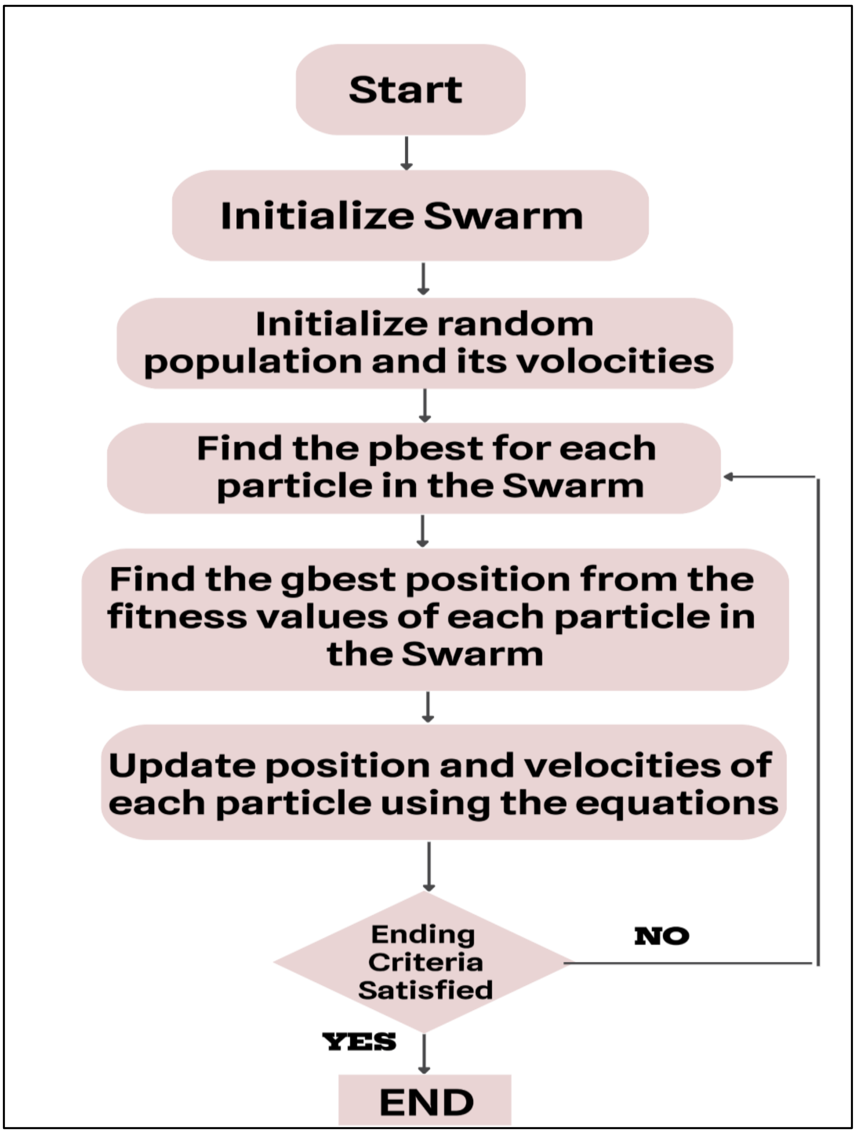

Figure 3-3: The Flowchart of PSO

- ***Initialization:*** The primary step is to generate a swarm of particles that creates a team uniquely representing each hyperparameter combination. Each of the candidate solutions is assigned positions and velocities randomly, ensuring good coverage of the search space.
- ***Evaluation:*** The fitness of each particle is measured by the extent of the contribution of its respective hyperparameter configuration to the predictive performance of the model. The measure uses the particular characteristics of the task to ensure that only the most beneficial features contribute to the model accuracy.

- ***Updates:*** After the check is completed, every particle changes its position by referencing two indicators, the particle's personal best (pBest) and the best-known overall solution (gBest), every time. The optimization algorithm adjusts the particle's velocity according to the distance of the current position of the particle from these two extreme values; thus, the particle is moved in a new path in the search space. This is how the new trajectory towards better solutions becomes possible.
- ***Convergence Check:*** One of the most crucial elements in the PSO mechanism is the need for a good convergence check, without which it is impossible to know when the algorithm should be stopped. The termination condition may be determined by acquiring a set fitness value or by reaching the maximum number of iterations. If any of these conditions are fulfilled, the fitness evaluation process is again done for final verification.
- ***Output:*** Upon reaching convergence, the PSO will then output the global optimum. This signifies that it has determined the combination of the hyperparameters that gives the best performance throughout the optimization. The role played by these parameters is to increase the LSTM model's pattern recognition capability and improve the model's predictive performance in a useful way.

The process of optimization undergoes a trade-off between model complexity and robustness, where both are required for the development of high-quality machine-learning models. In other words, the fine line between the two ensures that LSTM is well-suited to real-time text prediction tasks. The process of optimization involves a careful balance between model complexity and robustness, both of which are essential for creating effective and efficient machine-learning models. Ultimately, this subtle balance ensures that the LSTM is exactly designed for the intricacies of real-time text prediction tasks.

2.2.3. Long Short-Term Memory

Long Short-Term Memory (LSTM) networks are a type of RNN ideal for keeping information for a long time in sequential data. This property is the primary value of LSTMs, making them highly useful in various types of text prediction, particularly in search engines. In this system, the LSTM networks create and learn from a series of words, or queries, to forecast the next most likely letter or word based on the context given by all previous inputs.

In most cases, traditional RNNs face the well-known vanishing gradient issue and get stuck at learning from vast data sequences. Whereas LSTMs address this general problem by means of memory cells and trust-structured gating units. These elements regulate the data flow to various places, which makes us aware that the model has preserved the most important long-term relationships. The ability to detect user intent in search engine queries becomes the main aspect of generating the right advice and achieving an enhanced experience.

The heart of LSTM networks lies in the LSTM units, which act as the building blocks. Each LSTM unit has three gating mechanisms, which are responsible for controlling the information flow.

- ***Forget Gate:*** This gate is responsible for evaluating the significance of the available information in the memory and, from that, determining what is less essential and hence can be "forgotten". A good example in the case of search queries is where people need to know that the background may no longer be valid.
- ***Input Gate:*** The purpose of the input gate is to maintain the memory cell updated with novel and essential pieces of data. This technique enables the model to adapt to different query patterns or user input methods; thus, it remains useful in forecasting the next terms.
- ***Output Gate:*** The gate selects what knowledge will be taken out of the memory cell to the next step in the time series that will be used in the model to predict the next word of the query. It, therefore, serves as a prediction of the model for the next word that is most probable in the query.

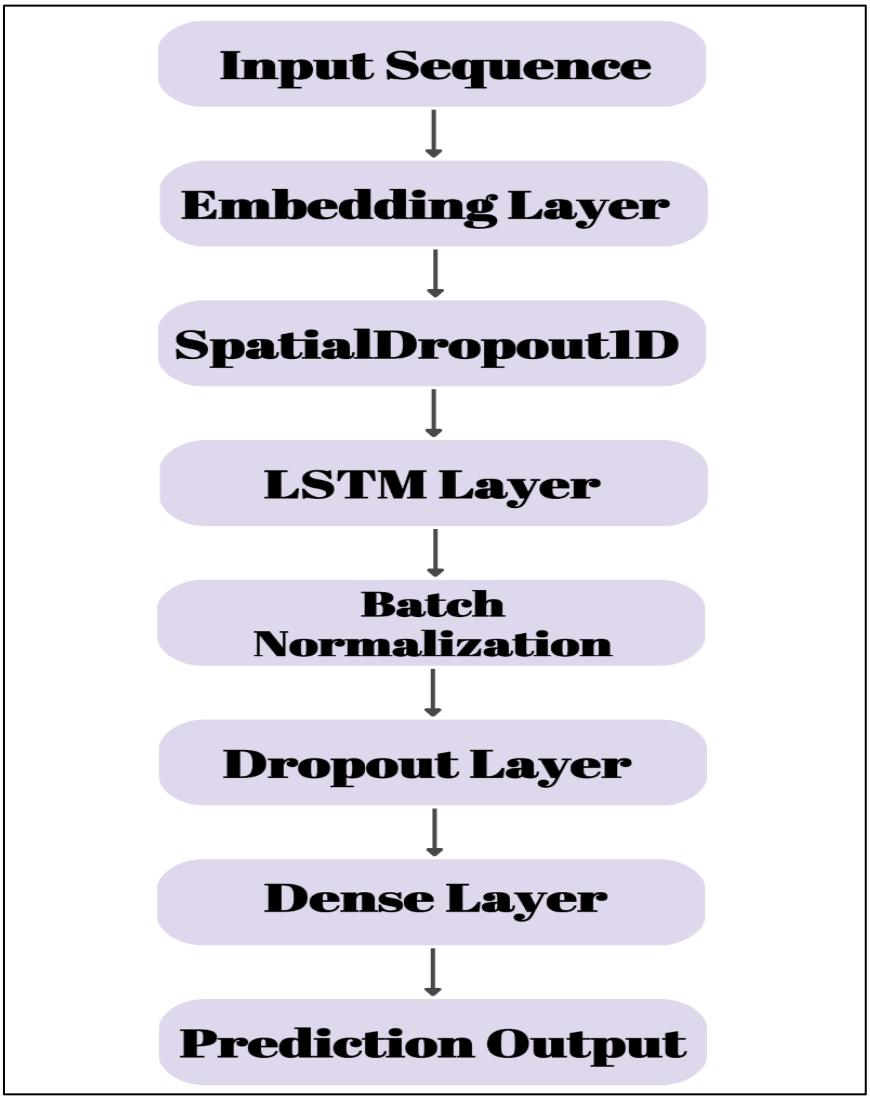

Figure 3-4: Flowchart of LSTM

Next, Figure 3-4 presents the well-built architecture together with the best hyperparameters to reinforce the model's ability to predict the text precisely, also having the advantage of low computation cost.

- ***Embedding Layer:*** This layer changes integer-encoded words into fixed-size dense vectors, referred to as embeddings. It simplifies the processing for the LSTM by converting the input text data into vectors, making it more efficient than traditional one-hot encoding methods. It gets the following parameters:
  - input_dim: States the total number of unique words in the vocabulary.
  - output_dim: Affirms the number of dimensions of the dense embedding vectors created.
  - input_length: Determines the length of the input sequences, filling the shortage with extra padding if required.

- ***SpatialDropout1D:*** It is applied to prevent the model from overfitting in the course of training by deleting some channels (1D feature maps) at random. This way, oversaturation of the channels can be well restricted, and the model will show better performance as there is no more redundancy. The parameters are as follows:
  - rate: The percentage of the input elements that should be discarded.
- ***LSTM Layer:*** This layer is in charge of processing sequence data to have a good hold of temporal dependencies. LSTMs, being powerful in terms of having long-term memory, can be applied in the realm of sequence data. The most important elements:
  - units: The number of LSTM units or memory cells set up.
  - return_sequences: Changed to False to notify that it is only the last output of the sequence that is required.
  - kernel_regularizer: Applying L2 regularization (e.g., l2(0.01)) is a step taken to solve the problem of the model overfitting. This is accomplished by discouraging the weights from growing too much in the process.
- ***Batch Normalization:*** This layer is useful as it changes the obtained results of the previous layer so that they get aligned to a 0 mean and a 1 standard deviation. By making such adjustments, this technique makes the learning process of the model stable and faster at the same time, which is achieved by the reduction of the internal covariate shift phenomenon.
- ***Dropout Layer:*** This level of regularization minimizes the risk of overfitting the model by excluding some of the input features randomly. By doing this, the model will not show only one unique representation, and in this way, an unconscious step in preventing overfitting is completed. These parameters are featured as:
  - rate: The fraction of input units to drop.
- ***Dense Layer:*** A multilayer perceptron that has the output layer as its last layer is responsible for generating vectors of probabilities over the vocabulary. The softmax operation is used for activation, ensuring the outputs are valid for multi-class classification tasks. It comes with the parameters:
  - units: The number of classes or the dimension of the output space (i.e., the vocabulary size).

- activation: The softmax function is employed to empower the logits with the probability property.
- kernel_regularizer: A regularization function Applies L2 regularization (e.g., l2(0.01)) to prevent overfitting.

This illustration demonstrates how LSTM networks, implementing such a comprehensive architecture and tailored components, are capable of providing text prediction services that are economical and intelligent in search engines.

2.2.4. LSTM/PSO Configurations

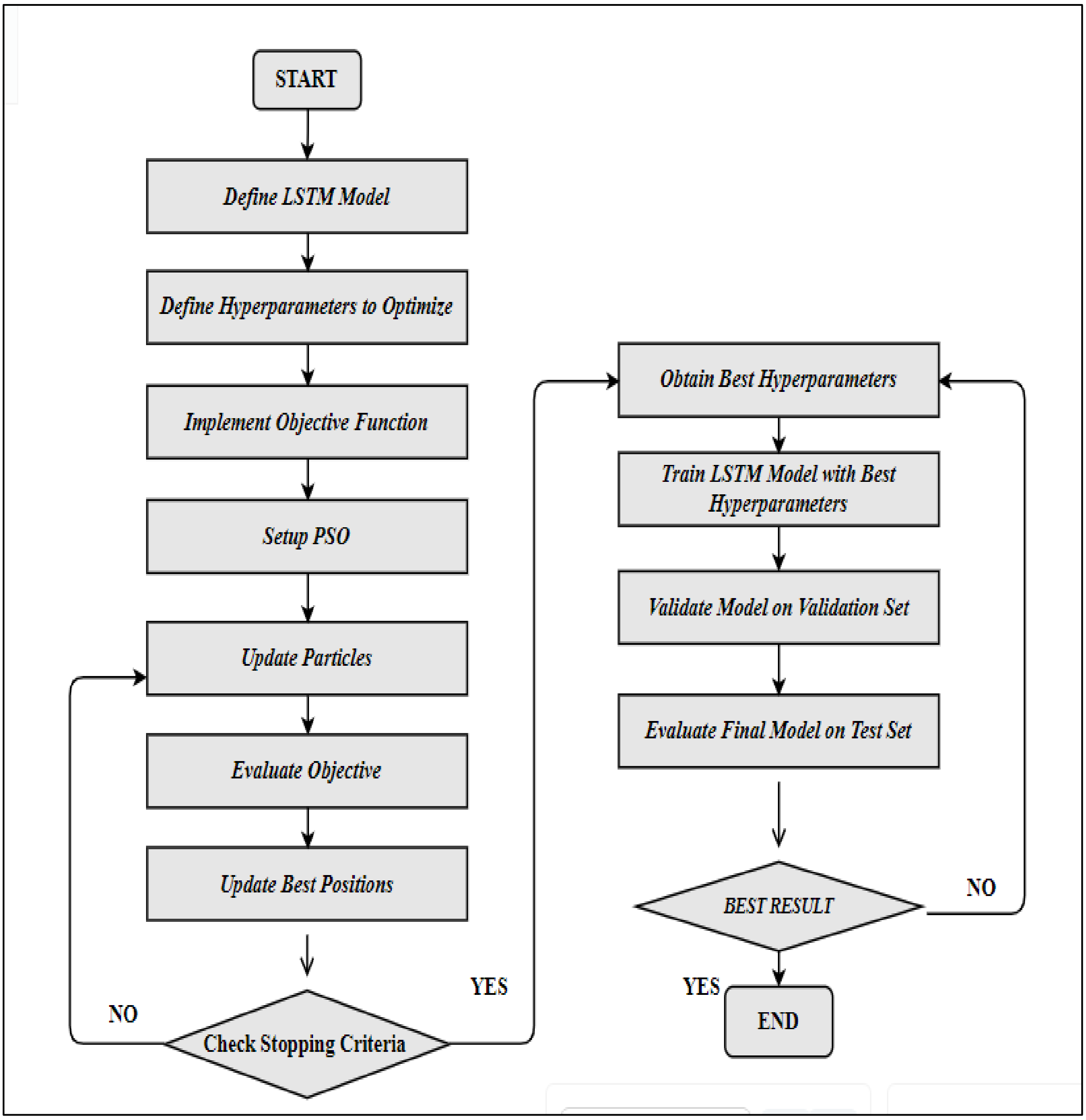

Figure 3-5: Flowchart of our LSTM/PSO Model

Figure 3-5 shows the structure of our PSO/LSTM model in the form of a flowchart.

- ***LSTM Model Definition:*** Provides the design of the LSTM network, indicating the number of layers, the hidden units per layer, and activation functions. Apply layer normalization and batch normalization for better stabilization and convergence.
- ***Hyperparameter Optimization Definition:*** Define the main hyperparameters in PSO.
- ***Objective Function Development:*** The objective function calculates the validation loss that the PSO algorithm uses to find the hyperparameters that minimize the loss.
- ***Initialization of PSO:*** Hyperparameter sets are represented by particles in a swarm whose positions and velocities are random.

$$V_{i,d}^{(t+1)} = \omega V_{i,d}^{(t)} + c_1 r_1\left(P_{i,d} - X_{i,d}^{(t)}\right) + c_2 r_2\left(G_d - X_{i,d}^{(t)}\right)$$

Equation 1: Initialization of PSO

- ω is the **inertia weight**, controlling exploration vs. exploitation.
- c_1, c_2 are **acceleration coefficients**, determining the influence of p-best and g-best.
- r_1, r_2 are random numbers drawn from U(0,1)to introduce stochastic behavior.
- $X_{i,d}^{(t)}$ is the current position of particle *i* in dimension *d*.
- $V_{i,d}^{(t)}$ is the current velocity of particle *i* in dimension *d*.
- $P_{i,d}$ is the best position found by particle *i*.
- G_d is the global best position.

- ***PSO Implementation:*** *For every iteration, the following steps are taken:*
  - *Particle Update:* Particles perceive their best-achieved position and global best swarm position, and thus, they alternately move around the hyperparameter space.

$$X_{i,d}^{(t+1)} = X_{i,d}^{(t)} + V_{i,d}^{(t+1)}$$

Equation 2: Particle Update

  - *Objective Function Evaluation:* Estimate the loss (the difference between the outcomes of the model and the real values) for each particle's hyperparameters.

$$f(X_i) = Loss\left(Model(X_i)\right)$$

Equation 3: Objective Function Evaluation

- *Best Positions Update:* Attach the best hyperparameter configuration found so far to the record from the statistics.
    - Update **personal best** if the new position improves performance:

$$P_i = X_i^{(t+1)} \text{ if } f(X_i^{(t+1)}) < f(P_i)$$

Equation 4: Personal Best Update

    - Update **global best** if the new best position is better than the current global best:

$$G = X_i^{(t+1)} \text{ if } f(X_i^{(t+1)}) < f(G)$$

Equation 5: Global Best Update

- *Stop Criterion Checking:* Define the termination criterion of the maximum number of iterations when either the necessary conditions are satisfied or the system cannot find the right hyperparameters.

- ***Acquisition of Optimal Hyperparameters:*** *The group with the minimum validation error post PSO runs is the one to be selected.*
- ***Training with Optimal Hyperparameters:*** Using the PSO received hyperparameters, conduct an early stop, and adapt the dynamic learning rate while training the LSTM model.
- ***Model Assessment on Validation Set:*** A means of evaluating a model is through a validation set using cross-validation, so that robustness is ensured.
- ***Final Model Assessment on Test Set:*** The nature of the model is investigated by a test set to be sure that it is accurate, namely by accuracy, precision, recall, and F1-score.
- ***Best Result Decision:*** Once the model is good enough, it is deployed, or else, check the situation of hyperparameters and training to get better results is recommended.

Table 3-1 is a pseudocode that depicts the steps to optimize an LSTM model by implementing the PSO algorithm.

Table 3-1: Pseudocode of the PSO/LSTM configuration

	Algorithm 1 PSO/LSTM pseudocode.
1.	**Initialize swarm_size, max_iterations, inertia_weight (ω), acceleration coefficients (c1, c2)**

```
Initialize swarm of particles with random hyperparameter sets
Initialize velocity of particles randomly
Initialize personal_best (p_best) and global_best (g_best)
For t in range(1, max_iterations):
  For each particle i in the swarm:
      Update velocity using:
          velocity[i] = ω * velocity[i] + c1 * random() * (p_best[i] - position[i])
                              + c2 * random() * (g_best - position[i])
      Update position using:
          position[i] = position[i] + velocity[i]
      Clip position values to allowed hyperparameter range
      loss = Train_LSTM(position[i])
      If loss < p_best_loss[i]:
          p_best[i] = position[i]
          p_best_loss[i] = loss
      If loss < g_best_loss:
          g_best = position[i]
          g_best_loss = loss
  If stopping_condition_met():
      Break  # Stop if convergence or max iterations reached
final_model = Train_LSTM(g_best)
Evaluate final_model on test data
Return final_model, g_best, test accuracy/loss
```

The LSTMs have their unique advantages in learning from repetitive user inputs. Nevertheless, success can be largely dependent on how well you deal with parallel task submission and the right selection of hyperparameters like learning rate, dropout rate, and the number of LSTM units. It is along this line that the PSO works to solve a problem, which is resizing a model, and, therefore, enables the quest for an accuracy-computational resources trade-off.

2.2.5. Implementation Details

We integrated PSO to optimize hyperparameters and train the LSTM model simultaneously. The primary focus of this technique is to specify the hyperparameter settings that are the key to the highest quality of the model's text prediction in the open classroom search engine.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 203, 100)          1995100

 spatial_dropout1d (Spatial  (None, 203, 100)          0
 Dropout1D)

 lstm (LSTM)                 (None, 32)                17024

 batch_normalization (Batch  (None, 32)                128
 Normalization)

 dropout (Dropout)           (None, 32)                0

 dense (Dense)               (None, 19951)             658383

=================================================================
```

Figure 3-6: the Architecture of LSTM hyperparameters/ layers

The model in Figure 3-6 is based on standard LSTM components and uses the PSO optimization method to determine the best parameters of the LSTM. The architecture begins with an Embedding Layer, which converts input sequences into a more concentrated, vector form. Next comes the Spatial Dropout1D layer at a rate of 0.5, which is there to battle overfitting by dropping the whole 1D feature maps during training. The third section of the architecture is the LSTM layer, which is here equipped with 32 filters and a regularization rate of 0.01, a measure that mitigates the risk of overfitting by discouraging overly large weight values.

Additionally, our network configuration utilizes both Dropout and Batch Normalization. The Dropout layer uses a retention rate of 0.5 and is immediately followed by a Dense Layer, which is activated using Softmax, to achieve a multi-class classification effect with an additional L2 regularization term of 0.01 to ensure model robustness. The next step is the implementation of mini-batch gradient descent with a batch size of 128, achieving more computationally efficient handling of the data, and the learning process has an optimal trend toward the minimum of the loss function. As part of our loss function evaluation, we used the "categorical_crossentropy" loss function, which is a model of choice for effectively solving a multi-class problem with string type encoding as the target variable. The optimizer used in the training is the popular "Adam" optimizer, recognized for its ability to adjust the learning rate dynamically.

The PSO configuration is a 20-particle swarm reconciling efficiency in exploratory and convergence tasks, and will undergo 30 iterations to search the complete solution space. The constant inertia weight value of 0.5 in the updating scheme makes PSO stable and also helps keep the selective pressure necessary to maintain the global and local populations. By setting the cognitive and social components to 1.5, the Agents will be provided with supportive additions required to achieve their personal best and that of their neighbors. The objective of the fitness function is to minimize the validation loss calculated on held-out data, thereby ensuring that the model is not only trained properly but is also able to generalize well to unseen data, increasing the model's performance and making it more reliable.

2.3. Experimental results

This section describes the experimental outcomes of the artificial intelligence-powered LSTM-PSO intelligent search engine, specifically designed for educational institutions to predict and adapt to changes quickly in the open classroom. Once the system construction and hyperparameter optimization processes were completed, the model was trained and tested with real-world educational data. The model's performance in the aspects of predictive accuracy, generalization capacity, training efficiency, and computation cost was investigated through several different experiments.

To ensure that the valuation is unbiased and can be replicated, the LSTM-PSO model was contrasted with various established deep learning models, including ordinary LSTM, BiLSTM, GRU, CNN, and the basic RNN. Such comparisons shed a bright light on the efficiency of Particle Swarm Optimization in the models that work sequentially. The performance measures, specifically, accuracy, precision, recall, f1-score, and perplexity, etc. are presented at different stages of the learning process to indicate the model’s progress and capability.

Along with numerical evaluation, visualizations of model accuracy and loss over training epochs are furnished. The diagrams illustrate the evolution of various aspects of the training process, such as converging, generalizing, and stabilizing the learning. Overall, these results not only prove the effectiveness of the suggested concept but also demonstrate its advantage in real-time, with human-like intelligence for educational purposes.

2.3.1. Quantitative Evaluation of System Performance

2.3.1.1. Performance Outcomes: Analytical Discussion

The inclines of the training and validation curves in the visualizations depict how the LSTM-PSO model works at various stages of training. They display both the model accuracy and the model loss over the epochs of three main training periods: 0-5, 0-50, 0-70, where training is accomplished.

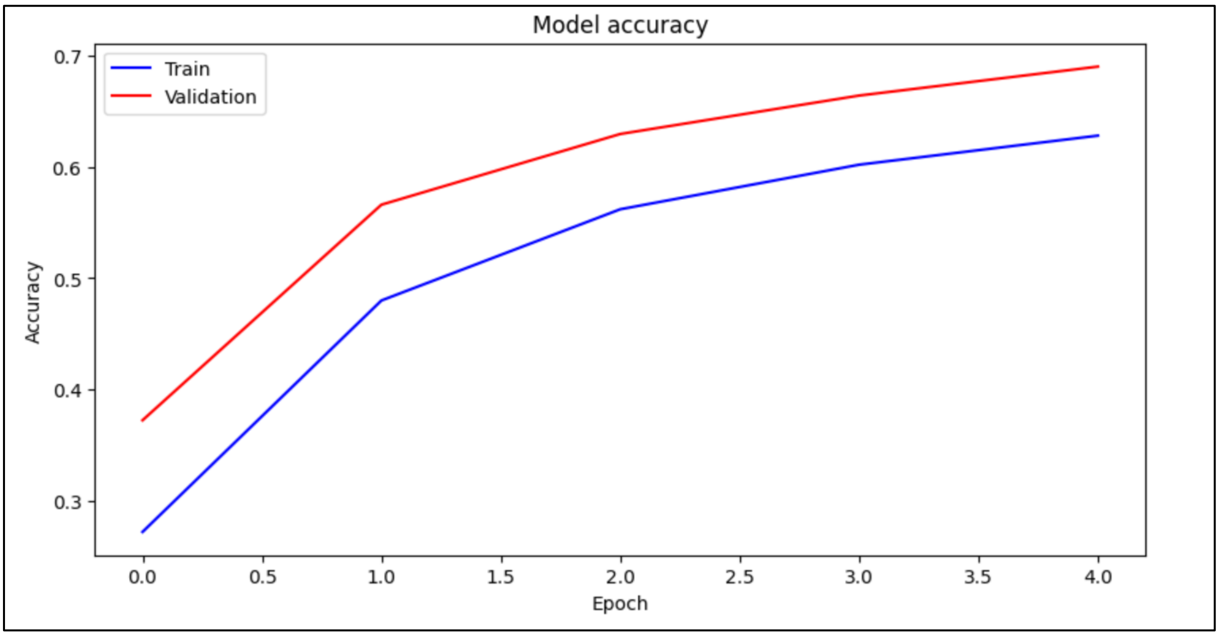

Figure 3-7: Train and Test Accuracy of the PSO/LSTM model at 5 Epochs

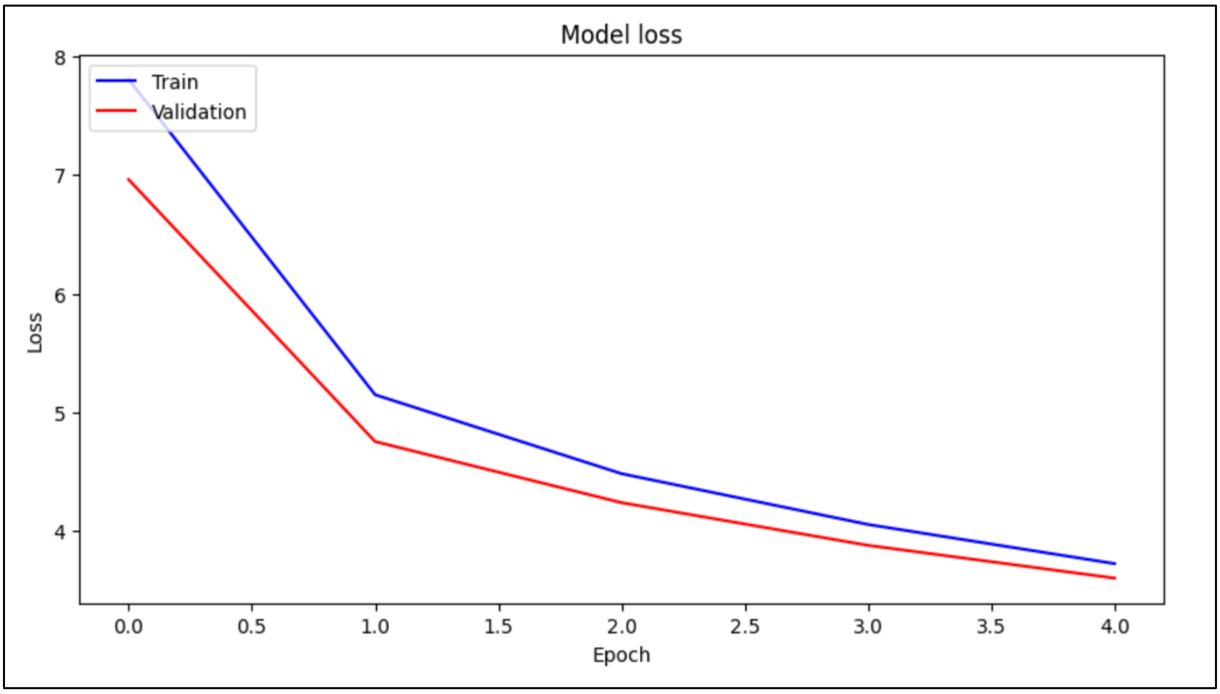

Figure 3-8: Train and Test Loss of the PSO/LSTM model at 5 Epochs

In the first five epochs of initial training presented in Figures 3-7 and 3-8, there is a considerable acceleration in the learning of the model, as illustrated by remarkably increasing accuracy and a significant decrease in loss for both the training and validation datasets. This initial part of the process is when the model begins to pick out important things among the data. The model can learn the basic structure of the data, but its accuracy is still less than 0.7, which is an insufficient level for any production deployment; hence, further training is required for it to be dependable in practice. Even when these earliest stages of training have been so promising as to indicate that a model can learn within the range, it is still not suitable for practical usage. Increasing the number of epochs is, hence, the most effective solution for further enhancing their performance.

From epoch 0 to 50, as illustrated in Figures 3-9 and 3-10, the training model's accuracy steadily improves, reaching a point where the training accuracy exceeds 0.8, while the validation accuracy is approximately 0.9 by the time it reaches epoch 50. In parallel, the loss curves also display a distinct flattening pattern, confirming that the model has stabilized in its learning. The small gap and the near-constant nature of the relation between the accuracy of the training and the test are the model's strengths. Hence, the chances of it being an overfit are minimized. This feature is seen as a stable performance on future data, which is one of the assumptions of reliable performance. During the 40-50 epochs, the model status becomes satisfactory. It's the right move to implement early stopping at this point, as it avoids overfitting and brings qualitative and quantitative performance.

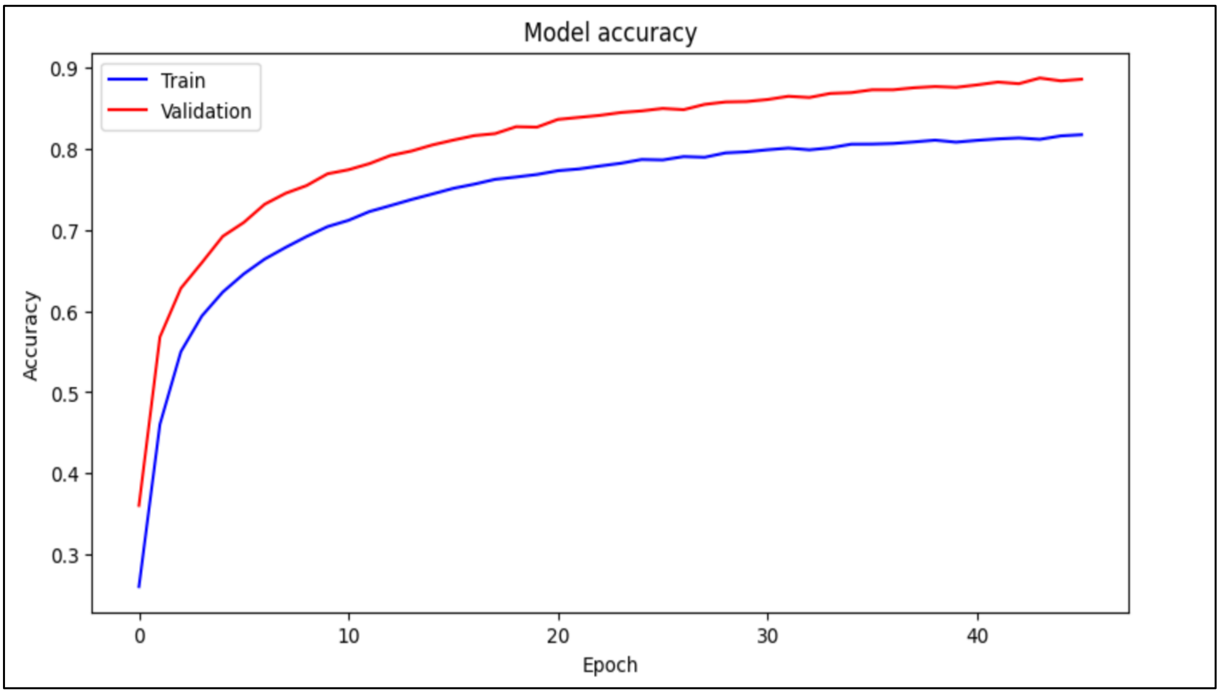

Figure 3-9: Train and Test Accuracy of the PSO/LSTM model at 50 Epochs

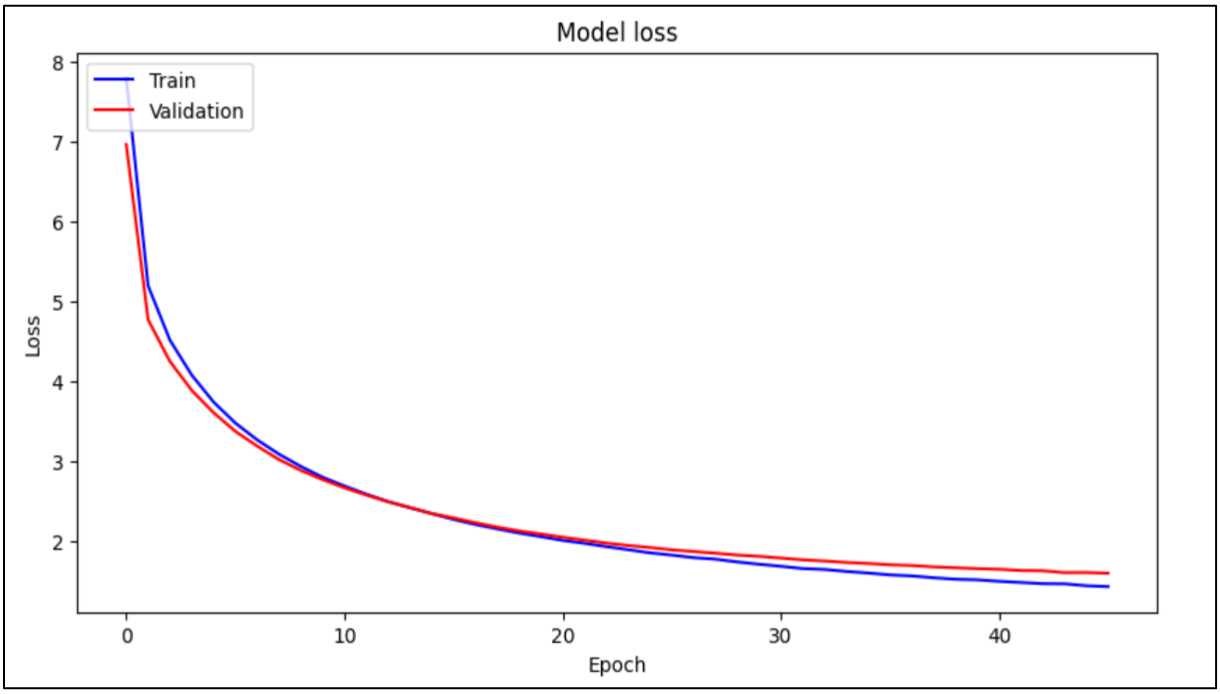

Figure 3-10: Train and Test Loss of the PSO/LSTM model at 50 Epochs

A conjoint analysis, which compares the accuracy and the loss plots at 70 epochs as demonstrated in Figures 3-11 and 3-12, has established that the LSTM-PSO hybrid model is not only viable but also capable. Throughout the training, the model demonstrated evident and continuous development of the performance indicators for both the validation and training datasets. Remarkably, the validation accuracy increased significantly to nearly 95%, and the training remained almost constant at approximately 86%. The visible discrepancy in accuracy reflects the strong generalization of the model, indicating that the model can not only very well understand the data pattern but also present new and good results when tested with unseen data.

Simultaneously, the loss curves exhibit a steep and continuous decline, mostly during the first 10 to 15 epochs, with the loss function decreasing from over 7 to about 1.5 by the 70th epoch. The significant plunge indicates the successful learning process. Likewise, the parallel course of training and validation loss curves throughout the training period is another piece of evidence of the model's regularization, which is configured well and dynamically responds to new data.

Additionally, the gradual flattening of the accuracy curve as well as the loss curve at the 55th epoch is an indication that the model has entered a state of convergence. The observation behind this is that if more training sessions are held beyond this point, the impact on the model's performance will be little, if any. Subsequently, the fact that the achievement of early stopping rules between the periods of 60 and 70 epochs brings about the availability of computational resources and the need to preserve the model’s quality is undeniable.

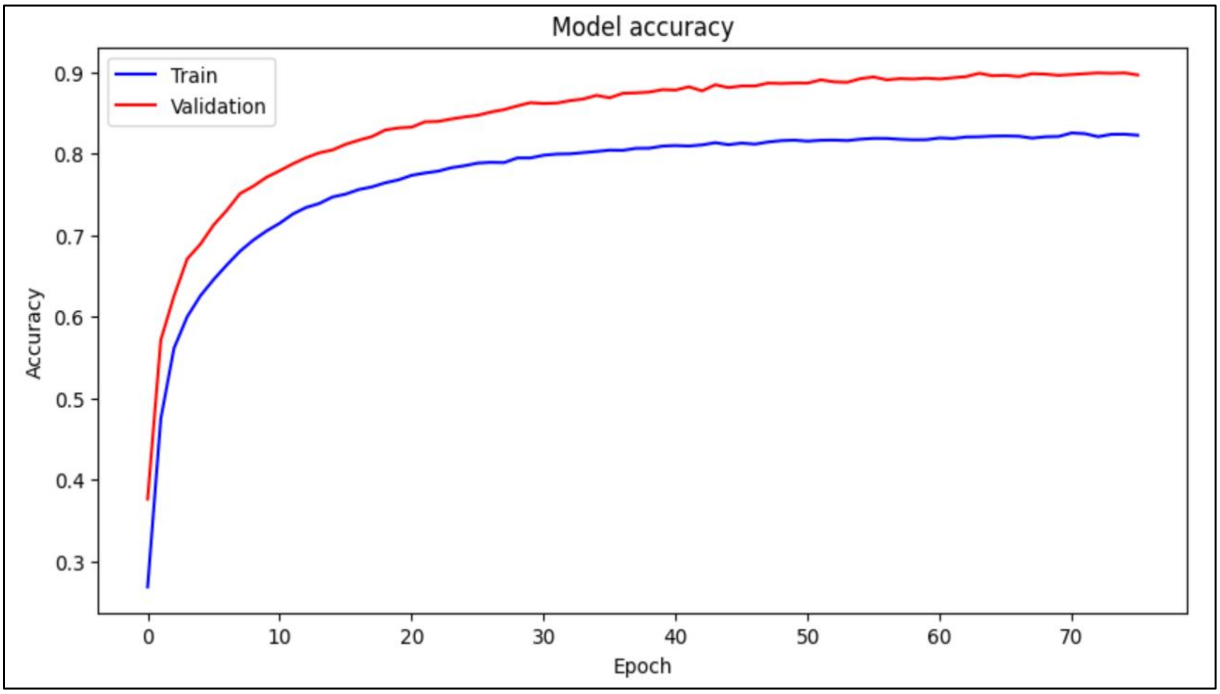

Figure 3-11: Train and Test Accuracy of the PSO/LSTM model at 70 Epochs

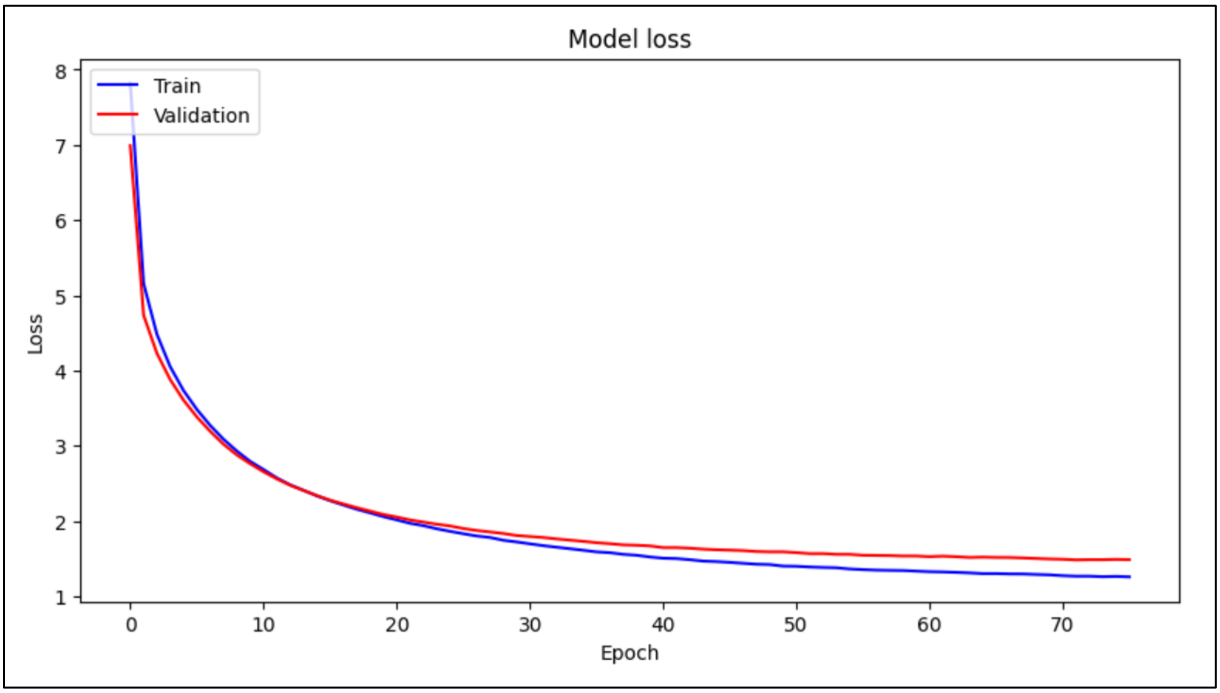

Figure 3-12: Train and Test Loss of the PSO/LSTM model at 70 Epochs

To summarize, the results of the 70-epoch training indicate that the LSTM-PSO model is both robust and efficient in its operation. The examination demonstrates that after 70 iterations, the model not only achieves adequate accuracy but also exhibits stable convergence and a strong resistance to overfitting. These features make it a practical and efficient solution in the context of predictive educational texts concerning real-time demand on platforms like Open Classroom, which requires both instant response and accuracy.

2.3.1.2. Performance Metrics Evaluation

To fully showcase the efficiency of the proposed PSO-LSTM model, we provide a detailed analysis of the maximum performance measures represented in the accompanying bar chart. The metrics cited in the paper are accuracy, perplexity, precision, recall, F1-score, and latency. These are all the metrics that, when put together, help understand how the model works during inference. The mark reflects the full issue in a compact form, allowing the efficiency check of the model across its most important dimensions of predictive accuracy and operational efficiency in a fast way.

- ***Accuracy:*** In terms of search engines, accuracy is the proportion of words the system can predict correctly related to the user's initial submission for subsequent words. Having a good accuracy rating means that users receive more pertinent suggestions than they expect and can engage with therefore enhancing design usage and relevance of the usefulness of search engines for searching. However, using accuracy as the sole indicator of performance can be misleading, especially in situations where search queries resulting from disparate datasets are not equally represented in the form of probabilities. If performance is based on inaccurate measurements with nothing to compare it to, simply using accuracy or similar measures inaccurately provides a skewed explanation of the effectiveness of the model.
- ***Perplexity:*** Perplexity is a measurement of the model's ability to comprehend the structure of human language. A lower perplexity signifies that the system can understand the syntax, identify user intent, and raise the likelihood of relevant search words being included. A model with a high perplexity score may suggest unclear words that may frustrate users and cause them to search ineffectively. It is important for all search engines to operate with lower levels of perplexity so that they generate natural suggestions that can maintain user perseverance and intention.

- ***Precision:*** Precision is important and related as it measures if the suggestions given by the search engine are relevant to the user's input. Precision measures the rate of suggested completions that the user deems appropriate. A model with a relatively high precision drives out wrong or misleading predictions and also clearly marks those to ensure users aren't troubled with unaffected autocomplete options. When Precision is inadequate, users may find themselves deleting suggestions or trying to re-enter the same query, which reduces overall productivity in a search. A strong Precision measure allows the model to supply the most accurate and contextually relevant results promptly, enabling more efficient searches.
- ***Recall:*** Recall is an important metric for assessing whether the model identifies all relevant words and phrases that are consistent with the user's request. High recall enables a wide variance of useful query completions, thereby increasing the likelihood that the user finds what they are looking for. Of course, some trade-off may exist. While high recall means irrelevant suggestions may be proposed, a low recall means the user may miss valuable keywords and phrases altogether. Achieving an acceptable balance between precision and recall is key; if users are able to successfully find a mixture of meaningful suggestions, while also having some reasonable, even if they do not make sense, suggestions, this can provide a rich and meaningful experience to the user.
- ***Latency:*** Latency refers to the time it takes the search engine to generate text predictions from the moment when a user submits a (partial) query. Keeping latency low enough to ensure good functionality and speed to deliver results to users (if latency is high, it could undermine effective search predictability). Excessive latency could undermine your users' overall experience with search suggestions, leading to stalled workflows and decreasing engagement. The search prediction model is trained to provide text predictions across the means of an effective, ultimately instantaneous suggestion system, enhancing the search experience to be fast and fluid.
- ***Memory Consumption:*** Memory consumption is an important characteristic that describes the operational capacity of a text prediction system. In very large-scale search engines, memory consumption could be a debilitating constraint to scaling and search latency. Effective use of memory is essential so that the model can predict responses quickly, while maintaining that level of responsiveness when many queries

are competing. Ideally, the model is designed to utilize minimal high-end hardware to perform adequately, even when thousands of searches occur concurrently.

- ***Training Length:*** Training length indicates the time taken for a model to learn from large data. It's significant for search engines to run models that can perform well with occurring changes in dynamic search trends, changing lexicons, and changing user behavior. Long train times can be dangerous; they could put the prediction systems at risk of being completely outdated and failing to serve the needs of the users. They can speed up the time taken to retrain a model significantly by using optimization techniques such as PSO, while creating a normal and reliable process to train that would enhance prediction systems to remain relevant and effective.

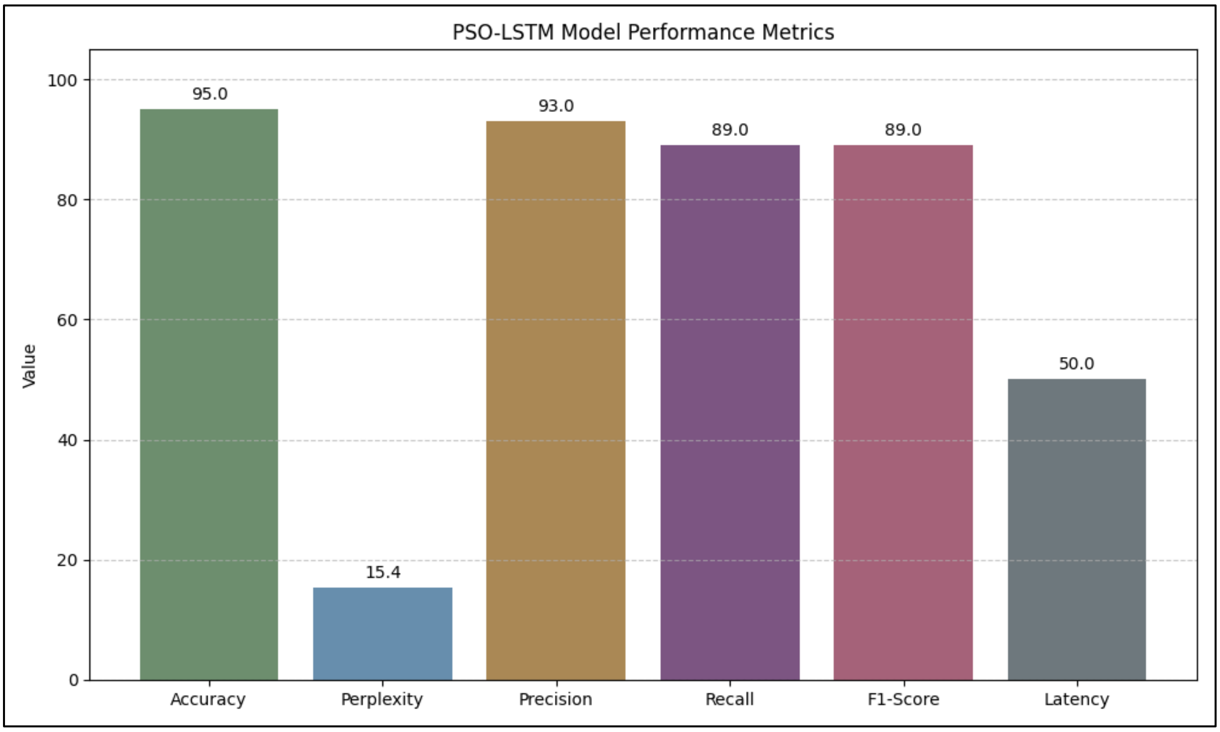

Figure 3-13: Key Performance Metrics of the PSO-LSTM Model

According to Figure 3-13, the PSO-LSTM model performs well with an accuracy rate of 95% meaning that the model is robustly predictive of the class. The model's precision of 93% reinforces the fact that it accurately outperformed the correct positive predictions. Concurrently, an 89% F1-score provides the impression of a perfect blend of precision and recall, thus demonstrating that the model can minimize neither false positives nor false negatives effectively; furthermore, the 89% recall rate indicates that the model is quite useful at spotting true positive cases. This means that the model is reliable for capturing consistent information that is crucial for the system.

The performance of the model significantly depended on the perplexity score of 15.4. This result demonstrates that the model is accurate and is almost absolutely confident in the predictions it is instructed to make. Hence, a low perplexity score not only indicates the understandability but also the tractability of the text that is being modeled in the task. Considering the system performance aspects, the model functions efficiently with an operational latency of only 50 milliseconds, which is necessary for real-time usage examples and smart educational platforms. This perfect match between the speed of making the right decisions and the waiting time for the output to be produced sets the PSO-LSTM a superior choice, as traditional models had to force a trade-off between speed and accuracy.

Overall, the bar chart certifies that the incorporation of PSO greatly increases the learning effectiveness of LSTM networks. It strengthens the PSO-LSTM model's position as a powerful and competent method for predictive text tasks, particularly in the dynamically changing educational sector. Thus, it becomes the main carrier of both accurate prediction and practical operational functionalities. Next, we propose an exhaustive comparative examination of our PSO-LSTM model against other popular architectures, underlining its relative advantages in real-world e-learning strategies.

2.3.2. Comparative Analysis with Existing Approaches

2.3.2.1. Cross-Method Performance Analysis

Figure 3-14 below clearly illustrates the various deep learning architectures varying in accuracy over 4 training checkpoints (5, 15, 40, and 70 epochs). This indicates their learning capabilities, generalization, and convergence – practical and helpful information. Whether from the charts or the depth reports, all model training arrangements were undertaken by us using the same pre-processed dataset. Therefore, the comparisons between models are fair, consistent, and controlled. The line trends in the chart evaluate and plainly indicate model maturation during training cycles, with a value towards decision making for choosing architectures under time-constrained or resource-constrained environments.

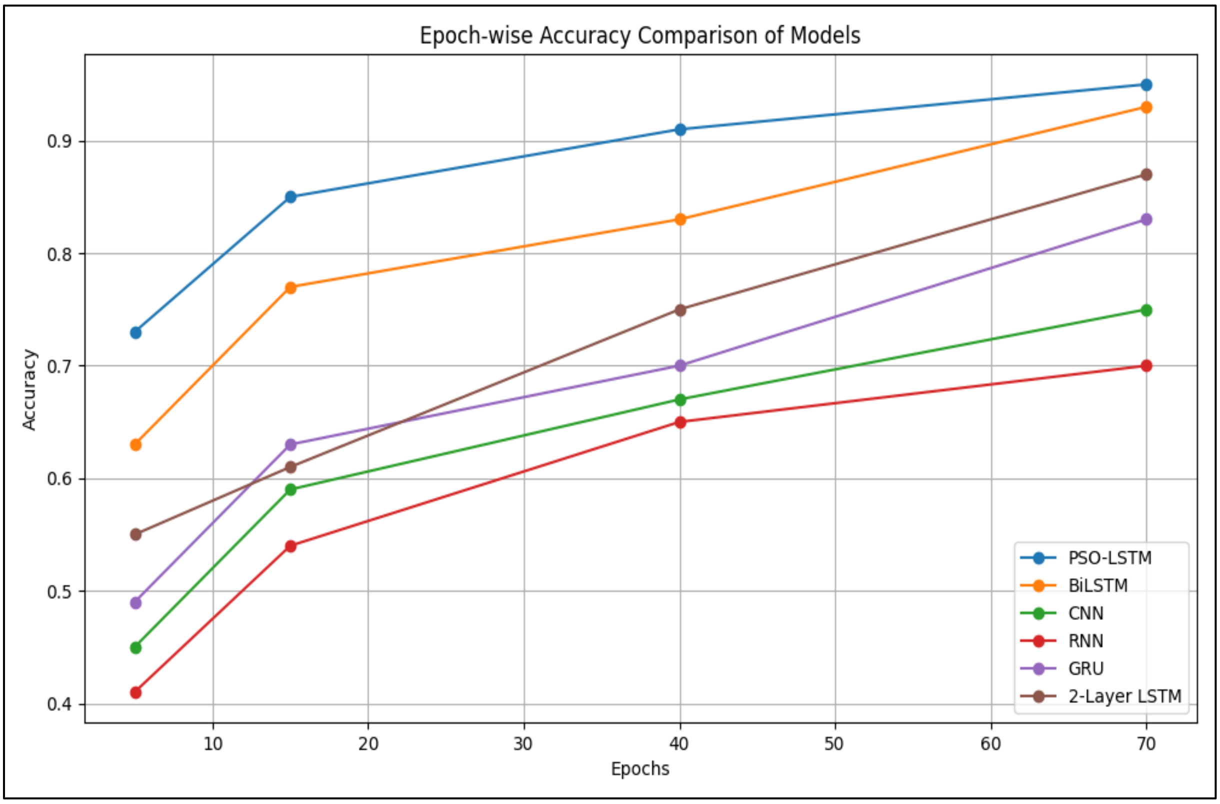

Figure 3-14: Performance Comparison with SOTA Methods

The PSO-LSTM hybrid model consistently exceeded the competitors throughout the process. It started at epoch 5 with an accuracy of 0.73 and had an impressive and effective increase towards the accuracy of 0.95 by epoch 70. The large learning curve, notably during epochs 5 - 40, highlights how PSO was a useful approach when optimizing LSTM hyperparameters. The advantage of PSO was automatic tuning, and therefore, the model was able to search through the hyperparameter space. The original experiment completed training in a shorter duration with better generalization ability. In addition, the performance curve of the PSO-LSTM remained relatively stable and substantiates its reliability and robustness from optimization-driven approaches.

The BiLSTM architecture shows a moderate but steadily improving learning curve, achieving an accuracy of 0.93 at epoch 70. Although it is close to PSO-LSTM in later epochs, it is slow to improve in early epochs. This is possibly due to the absence of automated optimization tools. This makes it sensitive to the nuances of manual hyperparameter tuning. BiLSTM benefits from bidirectional context modeling to further understand data sequences; however, this added complexity can be a contributing factor to slower convergence rates, particularly as hyperparameters might not be correctly tuned.

GRU presents a dynamic learning process, characterized by a steady growth in its accuracy from the starting point of 49% at 5 epochs to 83% at the 70th epoch. This kind of behavior of GRU reveals that this model is capable of learning rapidly and effectively, even when equipped with a small number of parameters, and still spearheads, which is a great advantage in situations where restrictions apply to memory and computational resources. The efficiency of the GRU model, in addition to speeding up the knowledge acquisition process, also enables a single-layer GRU to almost match the performance of the 2-layer LSTM by the 70th epoch. Its quick progress in the initial stages highlights its adaptability for tasks that demand fast deployment, therefore making this technology a good alternative in environments where efficiency, more than mere accuracy, is essential, with no accuracy sacrificed.

The CNN achieves a performance score of 75% at the 70th epoch, but its curve is much less steep in growth than the recurrent models. Starting from an accuracy of 45 %, the CNN algorithm shows regular advances but is visibly far from the performance capabilities of the LSTM-based model, as it is ill-suited for processing time-related problems inherent in such sequence data as texts. Although CNNs are strong in handling the spatial aspect of data and powering the tasks of image processing, they can be largely inefficient in discerning the linguistic aspects of such data, hence losing their effectiveness in particular areas that require a deep understanding of the order of events. This emphasizes the importance of carefully choosing model architectures based on the distinctive features of the data being analyzed.

The simple RNN is introduced as the model with the lowest performance among the models discussed, with an initial accuracy of 41% and then quickly reaching about 70%. This slowdown is mainly due to the basic nature of RNN, which deals in shorter loops, and therefore, it is almost impossible to capture the long-term relationships of the data. Additionally, they are notoriously vulnerable to vanishing gradient issues in the sequence of the data, making it hard for them to be efficient in learning. Although RNN is lightweight and straightforward to understand, it appears this network is still in the early phase of technology development that will handle sophisticated prediction tasks such as OCR tasks.

The 2-layer LSTM demonstrated the fewest advantages, having started with the lowest accuracy of 0.55 at epoch 5 and exhibiting the slowest improvement throughout training. However, even by epoch 70, it reached only 0.87 accuracy, which was the lowest accuracy level overall of all models explored in this analysis. The increased depth of the LSTM without optimization raises the potential for overfitting while contributing to a progressive decline in convergence rate if not strategically and intelligently tuned. The LSTM multi-layer dimensions should be considered if they are to yield advancements; however, the 2-layer LSTM, as it was implemented (regardless of epoch), will produce diminishing returns in accuracy performance.

2.3.2.2. Evaluation of Selected Methods Based on Performance Metrics

The bar chart displays the comparative efficacy of each model based on four measures of performance: accuracy, precision, recall, and F1-score, with each group of vertical bars representing a different model. The height of each bar in the group represents its score on the corresponding measure for that model.

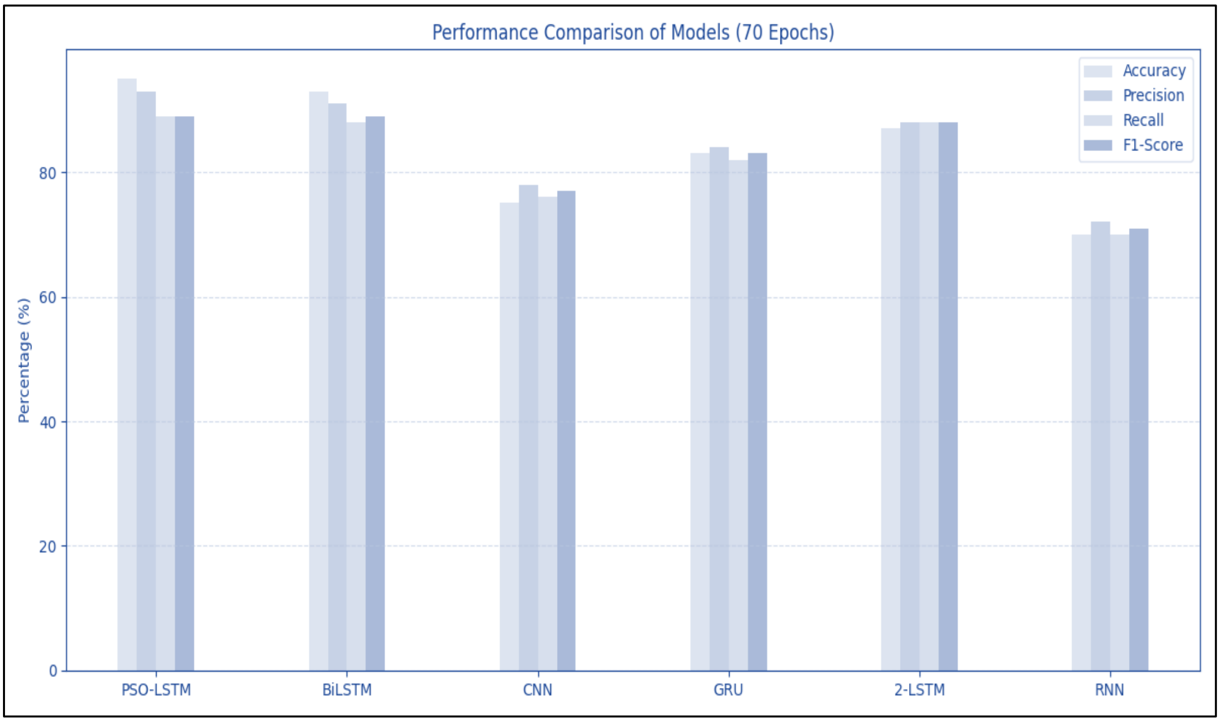

Figure 3-15: Performance Metrics Comparison

According to the performance summary in Figure 3-15, the PSO-LSTM model appears to be an effective choice, as it achieved a remarkable balance between accuracy, time, and resources. Furthermore, the PSO-LSTM has a solid chance of being one of the most competitive predictive analytics models existing. The perplexity score of 15.4 demonstrates a good trade-off concerning the complexity of the data, and a high accuracy of 95% clearly shows the usefulness of the stronger model. This high reliability suggests that it can be applied to a wide range of prediction problems for various industries.

The combination of PSO and LSTM has made this method viable for the automatic hyperparameter optimization step in model training, which not only reduces the time spent attempting to determine the best hyperparameter values, but PSO also avoids overfitting, a common challenge with complex models. As such, the PSO-LSTM provided exceptional and reliable performance, compared to the conventional LSTM and BiLSTM models, which are normally applicable to practical real-life problems.

In comparison, while the BiLSTM models deliver good accuracy averages of 93%, they require more latency, around 60 milliseconds, and heavier memory consumption, which peaked around 140 MB. Although they provided deployment opportunities across computational processes and public safety applications, they can significantly inhibit performance on real-time data processing and data retrieval. This reinforces the value of the PSO-LSTM model in these scenarios when speed and efficiency are critical.

Additionally, the 2-layer LSTM model competes well with the BiLSTM, though it, too, demands more time to train and requires a higher operational cost, making it less viable for commercial or larger applications. The current solution models of GRU and CNN can train with less energy costs (and resources), but lose even more performance gain rates, with 83% and 75% respectively. In this sense, these models can be utilized for applications where there are limits in computational resources and energy, particularly the GRU model at a relatively low cost. Receiving a Memory of 100 MB, it cannot only perform efficiently but is also more suited for deploying on edge devices.

The PSO-LSTM model combines practicality with accuracy, proving to be an excellent solution to be integrated into the search engines effectively. With a 50ms latency, it allows real-time applications that enhance the user experience. In addition to its memory efficiency of only 120 MB ensures adaptability to standard hardware and allows utilization without significant upgrades to infrastructure. With the optimization features of PSO, we can train a model without reducing time and resources for so long (just 5.6 hours); therefore, considerably sooner than BiLSTM and 2-layer LSTM. A key distinct feature of this model is its low perplexity value, which contributes to its accuracy in generating text and reducing errors. This is beneficial in areas such as query completion, chatbots, personalized recommendation engines, and particularly for a text prediction search engine that requires accuracy and prediction on user input without any lag.

Nonetheless, while PSO-LSTM has many benefits, it is essential to consider its limitations. First, although the PSO algorithm is invaluable, the overhead complexity that PSO requires needs to be considered to evaluate the time consumption during the hyperparameter optimization, which is a single investment but could be exhausting with limited resources. Second, the performance can be sensitive to hyperparameters, such as swarm size and inertia weight, and even potentially require more fine-tuning to achieve results. Third, while the PSO-LSTM system latencies and memory usage are acceptable for most real-time applications, systems requiring high throughput as a measure of performance may prefer the less complicated and more efficient models such as GRU or CNN, due to the more rapid inference times. However, although these simpler models are much more efficient than PSO-LSTM (for example, often running too slowly, not quickly generating predictions at times), they cannot match the context understanding and predictive performance of PSO-LSTM.

Table 3-2: Comparison between PSO/LSTM and other methods

N° Epochs	5	15	40	70
PSO/LSTM	0.73	0.85	0.91	0.95
BILSTM	0.63	0.77	0.83	0.93
CNN	0.45	0.59	0.67	0.75
RNN	0.41	0.54	0.65	0.70
GRU	0.49	0.63	0.70	0.83
2-Layers LSTM	0.55	0.61	0.75	0.87

Table 3-3: Performance Multi-Metric Evaluation

Model	Accuracy	Perplexity	Precision	Recall	F1-Score	Latency	Training Time	Memory (MB)
PSO-LSTM	95	15.4	93	89	89	50	5.6 hrs	120
BiLSTM	93	16.8	91	88	89	60	7 hrs	140
CNN	75	N/A	78	76	77	30	3 hrs	80
GRU	83	17.2	84	82	83	45	6 hrs	100
2-LSTM	87	16.5	88	88	88	70	8 hrs	160
RNN	70	21.5	72	70	71	90	4 hrs	60

2.4. *Discussion and Future Directions*

We have designed a real-time text prediction system that benefits from the integration of PSO with LSTM networks, for a search engine embedded within an OCR environment. In the fields of natural language processing, speech recognition, and time series forecasting, LSTM networks are often used for their capability of dealing with sequential data. For the data that the sequential nature is a feature of search engines, we envisioned an LSTM-based model to be the most suitable one for application in our intended use.

Among the reasons for increasing the efficiency of the LSTM model, PSO determines automatically the best hyperparameter settings. Inspired by the behavior of swarms of particles, it facilitates the exploration of the hyperparameter space. We have unveiled the PSO hybrid model capabilities to search for hyperparameters, making the learning process more efficient. We extensively compared our PSO/LSTM hybrid model with various machine learning models to demonstrate the abilities of the hybrid model and to prove that the approach proposed can consistently offer high accuracy results.

One of the important features of the PSO-LSTM hybrid method is its ability to capture long-term relationships in the time series dataset. Thus, the framework is excellent for applications in which maintaining context over an extended period is crucial, such as search engines. The major adaptation is made in traditional LSTM models - the PSO algorithm adapts the learning rate and batch size dynamically during training. Our experimental results demonstrate that the proposed model is capable of achieving a maximum accuracy as high as 95% with only 50 milliseconds of latency after extensive optimization of the loading time. This short latency has a significant role in being a differentiator, especially in applications where timely responses are essential, such as in real-time search contexts. This methodology demonstrates its value when dealing with complex and noisy datasets, an area where traditional models often remain unsatisfactory.

Aside from PSO and LSTM, we tested the use of the Bidirectional LSTM (BiLSTM) to investigate further their capability of input sequences in both forward and backward directions. This parallel processing enables the capturing of the context from around the words of the input sequence quite effectively. Here, the BiLSTM model is created to seek and carry on the relationships that words may have, even if they are in different parts of the sequence. Such an ability is useful in context-dependent applications, for instance, machine translation, where identifying the sense of other sentence parts is a precondition for generating reliable and natural-sounding translations. Although the BiLSTM model is quite remarkable in terms of performance metrics, having a precision rate of about 93%, one thing that should be pointed out is that this level of accuracy is only strictly achievable with a high computational cost. The complexity of BiLSTM models, a significant issue, may cause inefficiency in computational resource-limited settings, though the efficiency can be sufficient in general.

We have also explored CNNs, which have initially been used in image processing. CNNs employ convolutional filters to detect and analyze n-gram patterns in the data. Such an architecture enables CNNs to reapply the same structure in text analysis. CNNs, apart from their speed-up training process, also excel in feature extraction and pattern recognition. Nevertheless, there are certain tasks where they are challenged, for instance, in modeling sequential dependencies, which is crucial for providing a full understanding of the complexities of natural language, especially when dealing with long texts. The difficulty of the matter lies in the architectural characteristic of CNNs, which does not inherently allow the extraction of relationships on a broader level than the one defined by short phrases. CNNs have performed satisfactorily in text classification tasks, evidenced by a success rate of up to 80%. However, CNNs have been established to be less efficacious for sequential prediction tasks, such as next-word prediction or sentiment analysis in lengthy paragraphs, as they have trouble storing and relating information across long sequences.

RNNs are a subset of neural networks that are designed to handle sequential data. The difference between RNNs and traditional feedforward neural networks is that RNNs have connections that form a cycle; hence, they have memory. This architecture is what makes RNNs perform tasks such as language and speech understanding, financial time series prediction, and biological sequence analysis, to name a few. RNNs are very useful for dealing with sequences as they can process data one at a time, keep reminiscences about the previous inputs, correlate them with the new, and generate outputs. They are single when it comes to structure, and hence, they are easy to implement and train in quality. They not only extract temporality but also find wide application in problems like natural language processing or video analysis. However, traditional RNN models face the notorious vanishing gradient problem, which leads to a breakdown of their capability to capture long-term dependencies. This particular deficiency leads to a situation where the networks struggle to identify patterns that are spread over long distances. Thus, RNNs typically have an accuracy range of 70%-75%, which in some cases is sufficient, but one cannot compare it to the accurate results obtained with advanced architectures such as LSTM networks and GRUs, designed to address the limitations of RNNs.

GRUs are a streamlined architecture within the RNN family, simpler than LSTMs, as they have fewer gating mechanisms. It is clear that LSTMs excel in long dependencies with their complex structures, yet GRUs have the advantage of efficiently utilizing a single mechanism by merging the reset and update gates. This is energy efficient and saves the network the trouble of redundant gate storage. GRUs achieve convergence much earlier than LSTMs; hence, they require less training time, which is useful for large datasets. GRUs' compact architecture makes them particularly effective in demonstrating incredible performance, especially in scenarios where the probability of overfitting is very high and smaller datasets are involved. Generally, GRUs can reach performance similar to or better than LSTMs of about 80% to 83% accuracy. They are known to have fast training times and to maintain accuracy at highly impressive levels, which makes them robust choices for NLP (Natural Language Processing), time-series analysis, and speech recognition tasks as well. Essentially, GRUs are emerging into the deep learning community as the most sought-after method for providing a good balance between speed and accuracy.

A 2-layer LSTM network is a system with two levels of LSTM layers, which enables the model to recognize complex patterns easily from data. This technique permits a higher level of extraction, making it easy to extract the features from, for example, NLP and time-series data. The two-layered networks are generally better than one-layered ones, particularly in cases of complex tasks such as sentiment analysis or language translation, because they reveal more comprehensive patterns in the data and improve prediction accuracy. However, it's the same complexity that results in longer training times and higher memory usage, leaving the small-scale projects or those equipped with limited computational capacities with a challenge. For the 2-layer LSTM model compared to different LSTM architecture performance, a 2-layer LSTM model is usually found to outperform a single-layer model by 3% to 5% in terms of accuracy. Specifically, in the text prediction scenario, the most significant difference between the two is that the 2-layer LSTM is 87% to 88% precise on average. The higher precision of the former model denoted that it could adequately capture various complexities and intricate dependencies in textual data. Consequently, the 2-layer LSTM is an appropriate choice delivering impressive performance across numerous tasks, particularly in natural language processing and auto-generated text, where predicting precise sequence data is essential.

Table 3-4: Performance Summary Table

Model	Accuracy (%)	Training Time	Computational Cost	Best Use Case
PSO-LSTM	88–95	Moderate	Medium	Complex, noisy datasets
BiLSTM	85–93	High	High	Context-heavy tasks
CNN	75–80	Low	Low	Text classification tasks
RNN	70–75	Low	Low	Simple sequential tasks
GRU	80–83	Moderate	Medium	Resource-constrained systems
2-Layer LSTM	87–88	High	High	Deep sequential predictions

PSO has been our preference as it increases the accuracy and generalization of the model on the unseen data. LSTM has a complex structure due to its recursive nature and many hyperparameters, and thus is suitable for optimization with PSO algorithms. With a well-tuned LSTM, it is possible to make the model learn temporal dependencies more efficiently while avoiding common problems inherent in the vanishing gradient. LSTM/PSO hybrid approach is a strong contender due to the following factors:

- ***Optimizing Opportunities:*** PSO is speedy in finding the best configurations in a neural network transformation, such as learning rate and dropout rate, generalization of the model has been increased, and also the overfitting has been controlled.
- ***Sequential Data Management:*** LSTM excels at retaining data sequences, a fundamental point for working with natural language processing issues.
- ***Scalability:*** The model has the features of performing assignments of data effectively, which makes it suitable for services with heavier traffic, like search engines.

We demonstrated our smart text prediction system based on PSO-LSTM in this chapter. Such a system operates LSTM networks and PSO for hyperparameter tuning without human intervention. We identified our development environment, the preprocessing pipeline, and the model configuration, illustrating how we trained and tested a huge dataset. The visual instruction guided the readers through the system's operations. Examining visual results, it was clear that PSO-LSTM achieves the demanded performance level faster and without the pitfalls of other models. Overall, PSO-LSTM can be praised for its predictive ability and efficiency.

We aim to boost the resilience and flexibility of the PSO/LSTM model by adding new features. The first phase involves enhancing the PSO algorithm to respond to insights from both the learner and the teacher to adapt and achieve better results. Besides, it is planned to form different compound architectures, such as those of transformers and LSTM, to identify which exactly can provide a more accurate representation of the complex relationships in text data, and can lead to better contextual understanding, and on top of that, we will make use of our devised set of data.

3. Contribution 2: Integrating T5, Emperor Penguin Optimization, and IoT for Smart Open Classrooms: A Next-Generation AI-Driven Learning System

3.1. Description:

In this section, we will discuss the detailed implementation and testing of an AI-based virtual assistant (VA) system, specially developed for the OCR environment. In a bid to overcome some challenges existing by the traditional e-leaning systems like a monolithic uncharacteristic nature, the lack of personalization by the user, and non-contextual understanding of the student's environment, we come up with an inventive integration of three technologies, namely: the T5 (Text-to-Text Transfer Transformer), Emperor Penguin Optimization (EPO), and the Internet of Things (IoT).

The components integrate in perfect harmony to create an extremely clever, contextually sensitive, and hyperparameter-optimized VA that meets the diverse and evolving needs and changing learner requirements in a modern OCR environment.

Internally, the assistant utilizes T5 Transformer, a model known for its abilities in natural language understanding and generation, to make the communication between the user and the virtual learning environment more direct and interactive. It is essential to note that such an approach not only helps the assistant understand the most complex inquiries through the user but also generates the most coherent response that fits the context, thereby enhancing the learning process.

Furthermore, we utilize Emperor Penguin Optimization (EPO), an advanced hyperparameter tuning algorithm, and the best possible performance of machine learning models. With the implementation of this approach, we guarantee that the digital assistant remains effective not only in serving user needs but also in continuously evolving towards better-performing configurations, which in turn can offer personalized learning to individual students according to their specific needs.

With the help of IoT technology, this assistant becomes a data collector that can track in real-time changes in many aspects of the environment and users' attitudes. The assistant can thus change its recommendations immediately, if necessary, for example, due to changes in the setting or user preferences, while further engaging with the user and therefore allowing for a successful educational experience.

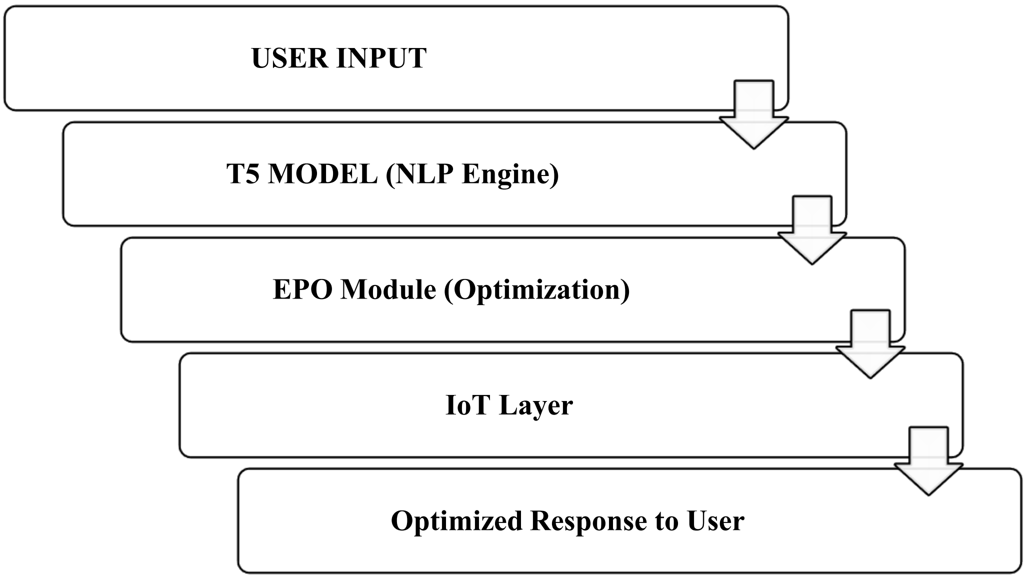

Figure 3-16: Overall System Workflow

In implementing this multi-dimensional strategy, our VA goes beyond the usual standards of responding, being accurate, and making learning support systems more personalized, thereby creating a more engaging and powerful learning environment that is in tune with the new era of learners' needs.

3.2. Methodology

This section presents a comprehensive study of the integrated technology of T5, EPO, and an IoT-empowered VA. Using these technologies, the system aims to achieve a user-friendly and interactive interface that employs T5 to implement natural language processing, with EPO for optimization tasks, and IoT for adjusting and responding to user necessities in real-time. The interaction of these systems is not only a more efficient way of working in real-time, but it also constitutes an attractive educational means of learning. The main ideas of this approach are the allocation of functional responsibilities and the characteristics of each element in the making of the ultimate VA.

3.2.1. Data Sources and Preprocessing

- ***Conversational AI datasets:*** In developing a chatbot specifically designed for the domain of OCR, we have resorted to the University Chatbot Dataset, which is available on Kaggle [160]. This rich dataset includes numerous pairs of questions and answers related to a wide variety of topics in a university, such as a guide to the admission process, the courses provided, the amenities of the campus, and the services offered to the students. The variety of this dataset makes it an ideal reference for building a dialogue system that focuses on the educational context. Moreover, we strengthened our model by integrating resources from multiple sources, such as scientific papers on conversational AI, user feedback from past bot chats, and industry trends in educational technology. This wide-ranging approach not only enriches the assistant's knowledge base but also ensures that it will continually be updated and be relevant and practical to use in the educational field. The preprocessing pipeline and the development environment remain consistent with what was outlined in the previous section.

- ***IoT Sensor Streams:*** This section focuses on capturing and compiling numerous environmental and biometric data streams. For instance, the recorded data samples include room temperature, the stress level of the people, noise, pollution, and other significant metrics that can reveal a great deal about environmental conditions and individual and collective health. The adherence of these various sensors enables the direct detection and analysis of the ongoing state, dispelling the myth of the possibility of unseen coming events and the necessary actions that can be taken. Preprocessing the IoT signals:
    - Outlier Detection and Noise Reduction: Brief abnormal spikes of the noise or heart rate signals were smoothed by applying a rolling median filter to reduce the influence of the motion artifacts and sensor mini-glitches.
    - Normalization: Each continuous variable was rescaled to the [0, 1] range to make all modalities have equal value and to avoid the algorithm being dominated by one variable.
- ***Embedding Creation:*** T5 models' embeddings are created to represent the semantic content of text. The design reimagines NLP tasks in the form of text-to-text, which helps in a better understanding of the context. With these embeddings, VA systems can get the most precise and semantically correct answer quickly, which is essential for keeping the system efficient and reducing consumers' waiting time.
- ***Data Augmentation:*** Data augmentation Implementation is one of the methods to ensure the model is exposed to diverse training data. The main methods of data

augmentation include paraphrasing, which is the process of using different words to explain the sentence with the same meaning, synonym replacement, replacing the exact words with their synonyms, and back-translation, which involves translating the standard language to a foreign language and then to the original language to generate a different sentence. These tactics together increase the diversity of the training set and help in generalizing as well as enabling the model to get used to various inputs.

3.2.2. T5 (Text-to-Text Transfer Transformer):

T5, also known as Text-To-Text Transfer Transformer, is a modern, advanced model in the domain of language processing that uses a transformer-based architecture to perform a wide range of language processing tasks. T5's most important feature is the method it uses to convert multiple NLP tasks into a standard text-to-text format. Such a systematic approach makes it convenient to solve tasks like translation, summarization, question answering, and many other NLP tasks with text as their only output.

The T5 model, proposed by Google developers, follows an encoder-decoder architecture, a key factor in both understanding the input language and generating coherent answers. The Encoder is of paramount importance in transforming raw text input into embeddings—complex, dense vectors that express a word's semantics and relationship to other words. To accomplish this, the Encoder proceeds to analyze in detail not only the word sequence but also the construction of the sentence and the contextual use of words. This feature enables the Encoder to obtain a proper symbolic representation of the input that can readily strengthen the model's comprehension in multiple tasks, such as text classification, sentiment analysis, and translation.

Given an input sequence X= $\{x_1, x_2,...,x_n\}$ the encoder maps it into a latent representation Z:

$$Z = f_{encoder}\,(X, \theta_{enc})$$

Equation 6: Latent Representation Equation

Where θ_{enc} Represents the encoder parameters.

The Decoder, however, forms the more intricate portion of the architecture; its primary job is to generate outputs that are not only understandable but also meaningful, from the embeddings provided by the Encoder. It employs attention mechanisms, which enable it to precisely locate and differentiate between the various contexts, moods, and intentions of the dialogue. The Decoder not only ensures that the responses delivered are logically consistent but also that they align with the nature and meaning of the conversation by investigating how the different linguistic elements of the input relate to each other. Identifying the critical aspects of the input that have crucial content has the advantage of giving the most accurate and appealing context-related information, which is also a part of the above mechanism.

The decoder then generates an output sequence $Y=\{y_1, y_2,...,y_m\}$

$$P(Y|X) = \prod_{t=1}^{m} P(y_t|Z, y < t; \theta_{dec})$$

Equation 7: Decoder Equation

Where θ_{dec} Are the decoder parameters.

T5 initially goes through a pre-training phase using a masked language modelling (MLM) objective. Firstly, the model is trained in a foundational manner, then proceeds to be fine-tuned. In the latter phase, the model becomes increasingly knowledgeable about the real meaning of user queries and the most reasonable suggestions in response to the user. The dual-stage method of training enables the model to be more efficient in reading the users' inputs and also producing the most appropriate answers according to the context for multiple NLP applications.

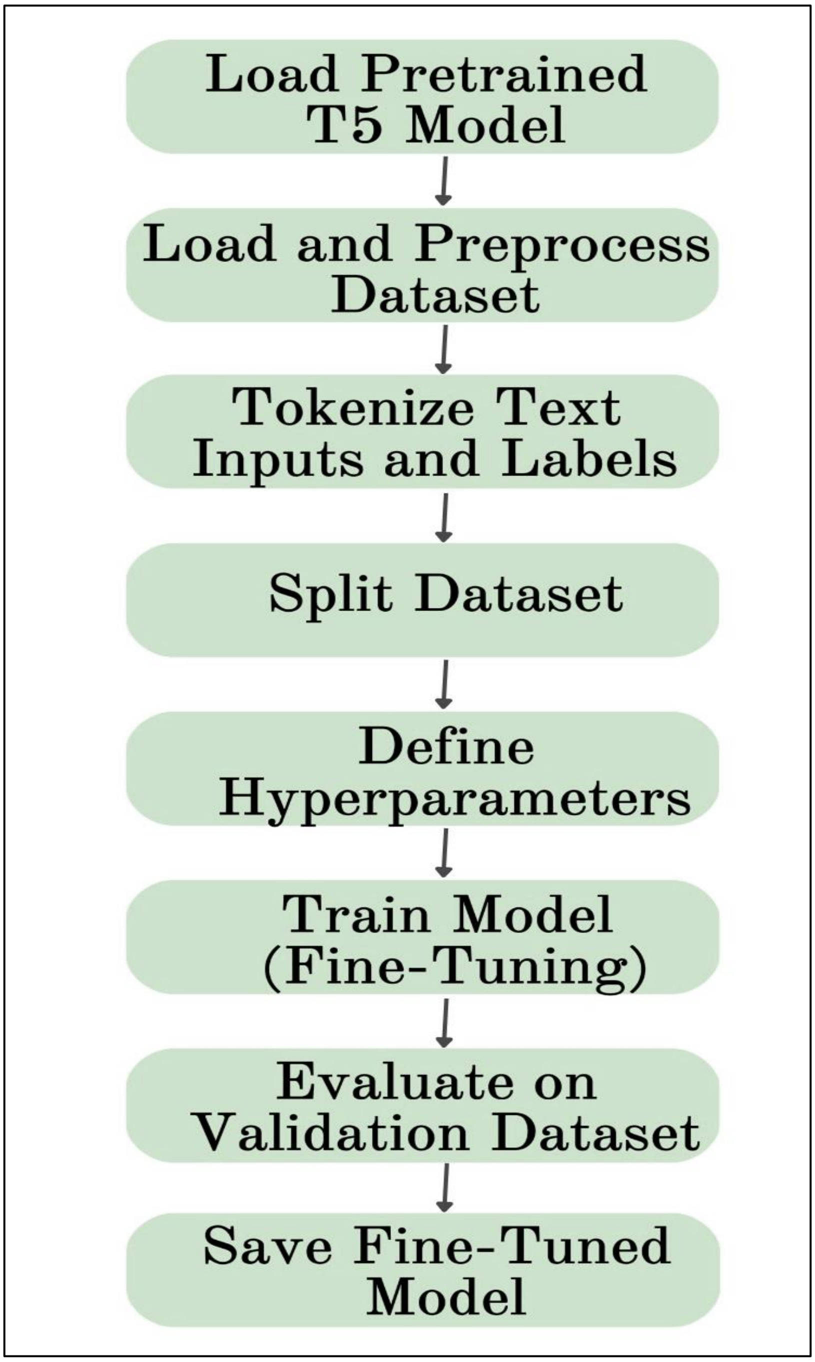

Figure 3-17: T5 Fine-Tuning Detailed Flowchart

The structured process of adapting a pre-trained T5 model for our VA is the primary focus of Figure 3-17.

- ***Loading the Pretrained T5 Model and Tokenizer***

The process of fine-tuning begins with the loading of the pre-trained T5 model and its corresponding tokenizer. These AI models have been trained on publicly available large-scale corpora such as Colossal Clean Crawled Corpus (C4), causing them to understand and generate human-like language. Utilizing Hugging Face's Transformers library, it is easier to import both the model and tokenizer into a programming environment, saving time and effort creating a new model. By following these steps, a well-established model is at hand, which itself has a sound understanding of language patterns and meaning, hence very complex these days.

- ***Preprocessing and Tokenization of the Dataset***

Once the pretrained model is ready, the next step is to import and prepare the data specified for the particular application, e.g., queries in the domain of education at OCR. Data preprocessing begins with text cleanup (by clearing the text of noise, special characters, or irrelevant tokens) followed by tokenization. Tokenization converts the text input into input IDs and attention masks using the T5 tokenizer. These new numeric forms are the ones that make it straightforward for the model to understand and process the sequence of the sentence while also indicating the sections of the input that are important for it to understand.

- ***Dataset Splitting***

The dataset is split into three parts: training, validation, and test. After the preprocessing, the training set that contains the largest portion of the data is used to update and fine-tune the model's weights; the validation set, utilizing a much smaller part of the data, is used to keep track of the model's performance, and finally, the test set carries the task of evaluating the model based on unseen data. This segmentation approach is indeed a check against simple pattern recognition of the input and an indication of robust learning. The validation is a good means to check whether the model has overfitted, and the decision could be made accordingly, when the training process has to stop, and the parameter has to be changed if done in the wrong way.

- ***Fine-Tuning and Loss Optimization***

Fine-tuning is the most crucial step in the T5 model, where the model learns the characteristics of the current task and adjusts itself accordingly. The trained model uses the labeled dataset to figure out which input queries correspond to the correct answers the model should provide. The loss or the deviation of predicted results from the correct answers in the task of training the model is constantly reduced by the loss function (in this case, the cross-entropy loss). The optimizer used is ADAMW, a revolutionary deviation particularly significant for transformer models, and an adaptive learning rate scheduler follows the optimization process. This action assures that the model acquires specialized skills in understanding and reacting to the specific type of task.

- ***Evaluation on the Validation Set***

During the fine-tuning process, the model is periodically evaluated on the validation set to observe the learning curve and identify problems like overfitting or underfitting. Thus, we track loss value, BLEU score, which is used to calculate the similarity between the response given by the model and the ground-truth response, and ROUGE score, which is applied to check the recall of key phrases. These indicators help us understand the degree of the model's capability to create suitable and logical results.

- ***Saving the Pretrained Model***

Once the model has received satisfactory training and successful validation, it is stored in association with a tokenizer. This preserved model is usable in real-life tasks like a Virtual Assistant for an educational platform. Keeping the model ensures that no retraining will be needed, and the processes are to be reused and scaled out across different systems. The latter is of high importance in the industry, where it is very costly to stop the production processes.

- ***Inference and Response Generation***

At the deployment stage, the newly trained fine-tuned T5 model is put into operation to process fresh user queries. To generate a response, when a student has a question, the input is first tokenized, then pushed through the model. The resulting output is next decoded into the desired human-readable language and is finally fed back to the user. It is now the inference step, the final point of testing the model's real-world utility, where it must be able to generate real-time, yet accurate, context-aware responses, thus confirming the entire success of the training pipeline.

3.2.3. Emperor Penguin Optimization (EPO)

The Emperor Penguin Optimization (EPO) algorithm is a state-of-the-art metaheuristic optimization method that mimics the amazing huddling behavior of emperor penguins, allowing them to survive the extreme conditions of their icy habitats. This new technique is based on the collaborative and adaptive principles of the natural world, with emperor penguins being the most prominent, as the birds primarily minimize heat loss and increase energy efficiency for their survival.

The essential feature of EPO is that the huddle restructures itself dynamically. In this paradigm, emperor penguins located at the edge of the huddle move inwards due to lower temperatures and accordingly save warmth; this happens completely naturally. Similarly, those located in the very center gradually move to the outskirts when they are too hot. Therefore, the thermal regulation is still in process, and it is more like an optimization process, where the goal is to obtain a perfect profile for every participant.

This algorithm encapsulates the mechanisms of a cooperative group of emperor penguins perfectly, wherein a two-pronged strategy enforces both exploration and exploitation. Exploration is about uncovering various potential solutions within a given search space, whereas exploitation extracts and refines those already identified. Due to this equilibrium, substantial enhancements in the highly important and complicated optimization problems are possible.

Interestingly, EPO has become the technology of choice in numerous technical areas, such as hyperparameter tuning in machine learning, where it is a significant asset, and the resource allocation and optimization areas, where it has shown positive outcomes. Through the intelligent and interactive behavior of these penguins, EPO provides a new path to solutions that are much more difficult to identify with traditional methods; at the same time, it is a stepping stone toward the development of machine learning and other optimization scenarios.

3.2.3.1. Hyperparameter Selection Process:

Our system includes the EPO method, which individually configures hyperparameters progressively to optimize the task. In order to do so, it eliminates the worst candidate solutions based on three critical criteria: accuracy, coherence, and efficiency. This policy employs a technical algorithm that automatically adjusts the hyperparameters with both reduction and increment actions.

It is through the initial evaluation of candidates' trials that EPO can identify the matching ones and then tune their hyperparameters to improve the system's performance. This process is continued until the best setup possible is achieved.

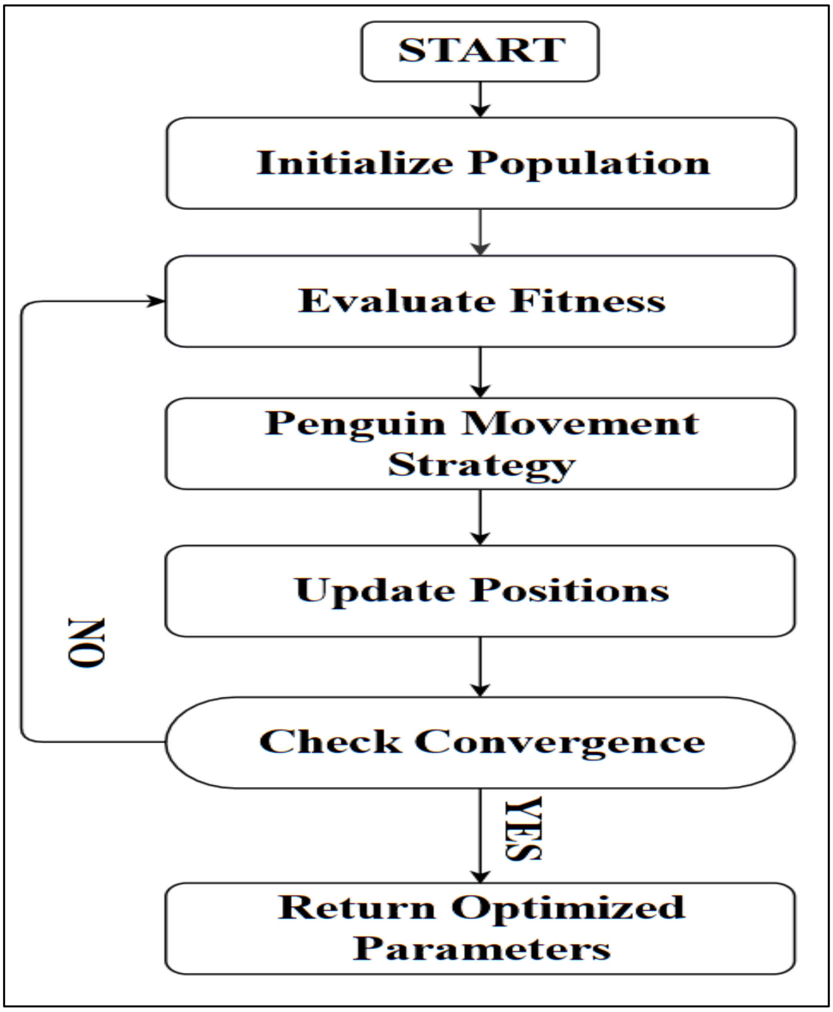

Figure 3-18: EPO Optimization in the T5-EPO-IoT system

Figure 3-18 illustrates the individual steps of the EPO optimization for the T5-EPO-IoT-Based VA. It demonstrates each operation in the optimization process and points out that the solutions were reviewed and refined.

- ***Population Initialization:***

In the beginning, a group of solutions will be created, and the population will be identified as $P = \{p_1, p_2, ..., p_n\}$. Where each solution, denoted as pi, refers to a possible solution within the specified area, thus a preliminary step for the following procedure.

- ***Huddling Effect & Movement Strategy:***

The movement strategy of the penguin inspired the algorithm to solve decision-making most efficiently, as mathematically formulated below by including parameters that are indicative of their huddling behavior. With this movement strategy, it can now explore the search space in a way that is both adaptive and directed, hence, leading to the algorithm being centered on local exploitation while maintaining collective welfare moving through the entire population.

$$X_i^{t+1} = X_i^t + S \cdot (X_{best} - X_i^t) + \alpha \cdot r$$

Equation 8: EPO Movement Strategy

Where:

- $\boldsymbol{X_i^t}$ Current position of the penguin (solution) at iteration t.
- $\boldsymbol{X_{best}}$ Best-known solution found so far.
- S Step size, adapting based on thermal balance.
- A Adaptive coefficient controlling exploration-exploitation tradeoff.
- R Random perturbation for solution diversity.

- ***Fitness Evaluation:***

The fitness value of each solution is determined through the use of an objective function that is specific to the goals of the optimization problem. In one example, for language-related tasks such as our Natural Language Processing (NLP) applications, it is the case that the underlying query is to maximize the BLEU / ROUGE score or the cross-entropy loss, illustrating that the generated text is of high quality and corresponds to the references.

Each candidate is evaluated based on:

- The amount of loss (cross-entropy loss).
- The degree of response correctness to get the BLEU and ROUGE scores.
- Resource management (speed and memory usage).

The fitness function is formulated as:

$$F(x) = \omega_1 . Accuracy - \omega_2 . Latency + \omega_3 . \text{resourceefficiency}$$

Equation 9: Fitness Function

Where $\omega_1, \omega_2, \omega_3$ are weight coefficients.

- ***Convergence Criteria:***

The optimization process continues iteratively, up until specific conditions that were determined in advance are found to be positive. These conditions include reaching the required point of accuracy, which indicates the model's performance, or an absence of any significant improvement over the past several iterations. It equates the function of the optimization process with the least number of resources.

After several iterations, the process converged, and the loss function stopped changing, indicating the best hyperparameters were discovered. The final setup not only lowered error rate in response generation but also made search results better, which made the virtual assistant more intelligent and better optimized.

3.2.4. Internet of Things (IoT):

The Internet of Things (IoT) is the fundamental pillar of personalized learning, enabling frequent real-time interactions and attaining engagement. It has the attribute of maintaining students' knowledge of the topics in a dynamic environment while integrating feedback, cognitive load, and learning modality tracking. In simpler terms, IoT adjusts learning materials and contents, attendance and schedule, equipment, and even the built environment according to individual learners' needs, thus providing them with the appropriate level of difficulty, length, and preferences.

One of the main features of IoT technologies is their integration with next-generation virtual assistants, which together offer a more personalized, efficient, and ultimately rewarding learning experience for students. In the case of a learner's system, IoT devices could play a vital role in allowing assistants to receive the human's capabilities and desires, thus, they can assist the people with the proper learning materials and provide them with personal advice. The platform also enables the use of language data collection that matches sophisticated NLP models such as the T5 model, which ultimately gives the students a more humane way of communicating and therefore increases engagement. At the same time, tools like EPO can utilize data analysis algorithms to make decisions on which learning path to set and which resources to allocate more productively, and rearrange classroom settings to ensure they are effective, as well.

IoT is indeed a major element in adaptive learning systems, which results in real-time interaction by observing users' behavior firsthand and analyzing their learning patterns. This combination also helps agents assist in a more customized and efficient manner in the education sector, thanks to the increasing use of highly advanced tools in NLP and optimization. Principal Functionalities of IoT within the System:

- ***User Insights:***

IoT products are constructed of sensors that can measure multiple biometric parameters, including stress, engagement, and physical posture. With the help of the collected data, the system understands the user's current health status, and thus is able to react interactively to progress the process of learning with a person.

- ***Environmental Background Adjustment:***

The adaptive technology automatically adjusts the OCR to be the best place for studying, not only concerning concentration but comfort as well. The most important environmental parameters, such as temperature, light, and noise, are under control and can be changed to those that benefit studying. Without a doubt, this smart setting will not only respond to poor user engagement by adding more light but also alter the temperature following the specifics of comfort recognition.

- ***Automation of Activities:***

Using the IoT system, the learning process becomes more efficient, i.e, by automating activities like pushing reminders, notifications, or alerts and providing intelligent guidance. Such smart tasks are made possible through the real-time data sources that are collected from IoT gadgets, and as a result, the users' experience becomes very rich, and the content is interesting to read.

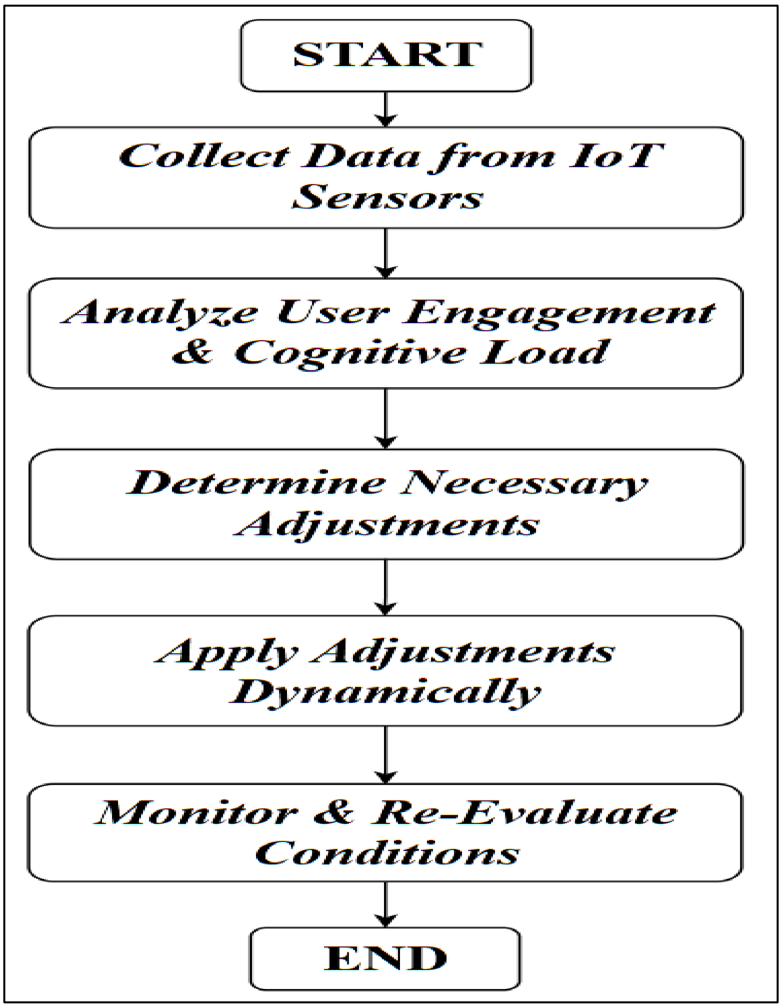

Figure 3-19: Flowchart of the IoT Functions in our System

The IoT enables real-time data acquisition via a chain of interconnected smart devices and sensors, which are always undergoing analysis and adjustments. The virtual assistant utilizes IoT to modify the learning environment intelligently. Figure 3-19 visualizes the concept of IoT in the virtual assistant system:

- ***IoT Devices Data Collection:***

This step consists of obtaining the necessary information for various sensors, such as wearables and intelligent lighting systems. From the data gathered, an image of the learner's condition at any given time is constructed. An example, the data includes cognitive load estimations, stress levels, user presence, lights adjusted to a certain level, temperature, and the voice activity for interaction.

- ***Data Processing and Analysis:***

- *Signal Processing:* This stage is about cleaning and analyzing the sensor data to find out which patterns and features are useful for the user's learning experience.
- *Context Awareness:* The model is designed to take into account the fluctuating environmental conditions and their effects on the learner's competence to make it capable of suggesting necessary alterations.
- *AI-driven Adaptation:* The system uses machine learning algorithms to change environmental settings, e.g., by dimming the lights, to prevent eye strain, after having analyzed the data, such as those that make sure the learning environment is ideal.

- ***Adaptive Learning Responses:***

  - *Adjustments within a Real-time Environment:* Various environmental conditions can be changed by the system on-the-fly, i.e., it could become lighter or darker, the screen displays fit or not to reduce mental load, and can be suitable for noise level management. It might also introduce relaxation techniques when rising stress levels are noticed as a means to uphold learners' welfare.
  - *Behavioral Insights for Continuous Optimization:* The significance of continuously following up and analyzing user activity data can't be ignored. All-time recording and inspecting users' logs allows the system to figure the crucial changes that might be necessary over time; thus, further experience can be gained and enjoyed.

- ***Stress and Cognitive Load Measurement***

Wearable IoT devices such as smartwatches, fitness bands, and posture monitors are the source of data to estimate users' cognitive and emotional states. These devices measure and record recent physiological signals of the body, such as heart rate, skin temperature, and movement patterns. Their signals are sent via Bluetooth or Wi-Fi to a local processing hub or cloud platform, which is a virtual assistant integration. Data preprocessing, normalization, and signal passing are all done in the signal analysis layer. These numbers serve as the input for the environmental adjustment algorithm, which is the component that enables the system to change its behavior instantly depending on the physiological state and the level of the learner's attention.

- To measure the stress levels precisely, the system uses various biometric data, which are easily accessible and are collected from wearable devices. Such data include:
- Heart rate readings are used to capture changes that might signal the occurrence of stress.
- Simple skin temperature sensors to pick up changes linked to the presence of physical or emotional reactions.
- Basic motion or posture data for further explanation of the primary data.

Working on the cognitive load side, the system implements easy-to-deploy instruments such as:

- Monitoring screen interaction patterns (e.g., time on task, number of clicks)
- Noting the pauses or hesitations during responses
- When available, an optional camera-based tool for assessing facial expression and engagement could be integrated.

Each of these sensor outputs is thoroughly normalized to a value between 0 and 1. This process of standardizing the varied data inputs ensures consistency in decision-making for educational adaptations.

- ***Environmental Adjustment Factor (Eadj)***

Using sensor data readings, the system design introduces a metric called the Environmental Adjustment Factor (E_adj), which is mathematically represented as:

$$E_{adj} = \sum_{i=1}^{n} \omega_i \, . S_i$$

Equation 10: Environmental Adjustment Factor

Where:

- Si is a normalized value of the reading from the i-th sensor.
- Wi is the weight given to the i-th sensor, showing how important that sensor is for the system's adaptation processes.
- n is the number of sensor channels that were analyzed in the calculation.

The weight coefficients Wi result from the initial system calibration, where they are empirically selected based on observed learning behaviors in various scenarios. For example, when cognitive overload is detected and indicated by an Environmental Adjustment Factor exceeding 0.7, the system triggers a dynamic modification of the user interface, which could include reducing task complexity, providing motivational prompts, or switching to a more visually engaging content format.

The E_{adj} calculated is sent directly to the real-time decision unit, which allows different modifications to improve the user experience. Such changes may involve not only altering user interface features like brightness, contrast, and color, but also changing the pace of instruction and pausing for a moment when too much information is given. Moreover, the mode of learning can be changed to go from reading to watching a video if the user is not interested. A control like this enables the virtual assistant to synchronize with the user's changing needs and preferences, in addition to addressing the content of the inquiry.

Within the T5-EPO-IoT framework, IoT sensors are crucial in custom-fitting learning material to the student's local physical and mental conditions. Through steady real-time biometric and environmental data collection, the sensors offer valuable information regarding crucial parameters like stress levels, cognitive load, and total engagement, which are essential for creating a tailored learning experience.

3.2.5. T5-EPO-IoT-based VA Configuration

Figure 3-20 illustrates how the T5 pre-trained model undergoes effective adaptation through the EPO method for hyperparameter tuning, developed for an IoT-driven environment and use-based knowledge OCR applications. The strategies used in the tuning and training process each time improve the versatility, accuracy, and reactivity of the virtual representative to the user's ease, more expeditiously.

- ***Load Pretrained T5 Model:***

Begin by importing the T5 model from the Transformers library created by Hugging Face. Initially, this basic model, pre-trained on diverse data, is to be refined and extended to conduct effective further adaptation.

- ***Load and Preprocess Dataset:***

The second step requires identifying a dataset that contains complementary educational text and IoT sensor data. This task involves a complete cleaning of text, the subtraction of noise and irrelevant information, normalization to ensure all data fits the same pattern, and the scaling of IoT data to get values within a certain range to have good model training.

- ***Tokenization:***

Use the T5 tokenizer to convert the preprocessed text and the IoT sensor into a form that can be fed into the model. This stage is important because it splits the text into tokens for the model's comprehension and further represents the IoT data for the content to be contextually rich.

- ***Split Dataset:***

It is necessary to divide the data into three parts: 80% for training, 10% for validation, and 10% for testing. This stratified method ensures that the model is not biased by the data's nature and will perform better when dealing with new data.

- ***Define Hyperparameters:***

Determine the set of important hyperparameters that have a huge influence on the performance of the model. These parameters consist of the learning rate (which lets us see how big of a step we're taking at each movement of optimization), the batch size (or the number of samples used in one iteration to calculate the error and then update the internal parameters of the model), the dropout rate (which is used to reduce the degree of overfitting that prevents the model from performing well, and it's implemented by randomly removing some of the neurons during the training stage), the weight decay (which is a regularization strategy by deciding the penalization factor), and IoT adaptation coefficients to make sure that the model has been suitably modified to incorporate the specific characteristics of IoT.

- ***Optimization of Hyperparameters using EPO:***

Use the EPO algorithm to reflect the performance of the chosen hyperparameters systematically. In brief, it appears to be ideal for efficiently finding an adequate model with a much higher level of performance in the described configurations.

- ***Train Model:***

Next, use the most refined backpropagation algorithms to enhance the model and merge them with the powerful AdamW optimizer to cope with sparse gradients during training and to incorporate the feature of weight decay. The IoT context is used for training to enhance the model's understanding and react to real-world sensor data.

- ***Evaluate on Validation Dataset:***

Once the model has been trained, its performance with the validation dataset is checked. This evaluation utilizes a few quantitative metrics, such as BLEU and ROUGE scores, to assess both the quality of text generation and the IoT compatibility, as well as the time taken to return the model's decision, thus the practicality of the model.

- ***Adjust Model Based on IoT Feedback:***

The model will change its responses dynamically to the actual IoT feedback. This capability of real-time adaptation proves to be necessary. Therefore, the model will be able to enhance customer satisfaction and user involvement.

- ***Save Fine-Tuned Model for Deployment:***

Finally, once the model has been trained and tested for performance and the validation is complete, it is stored. It is now that the model can be converted from the training environment to the deployment stage, and then it is usable in a real-life environment.

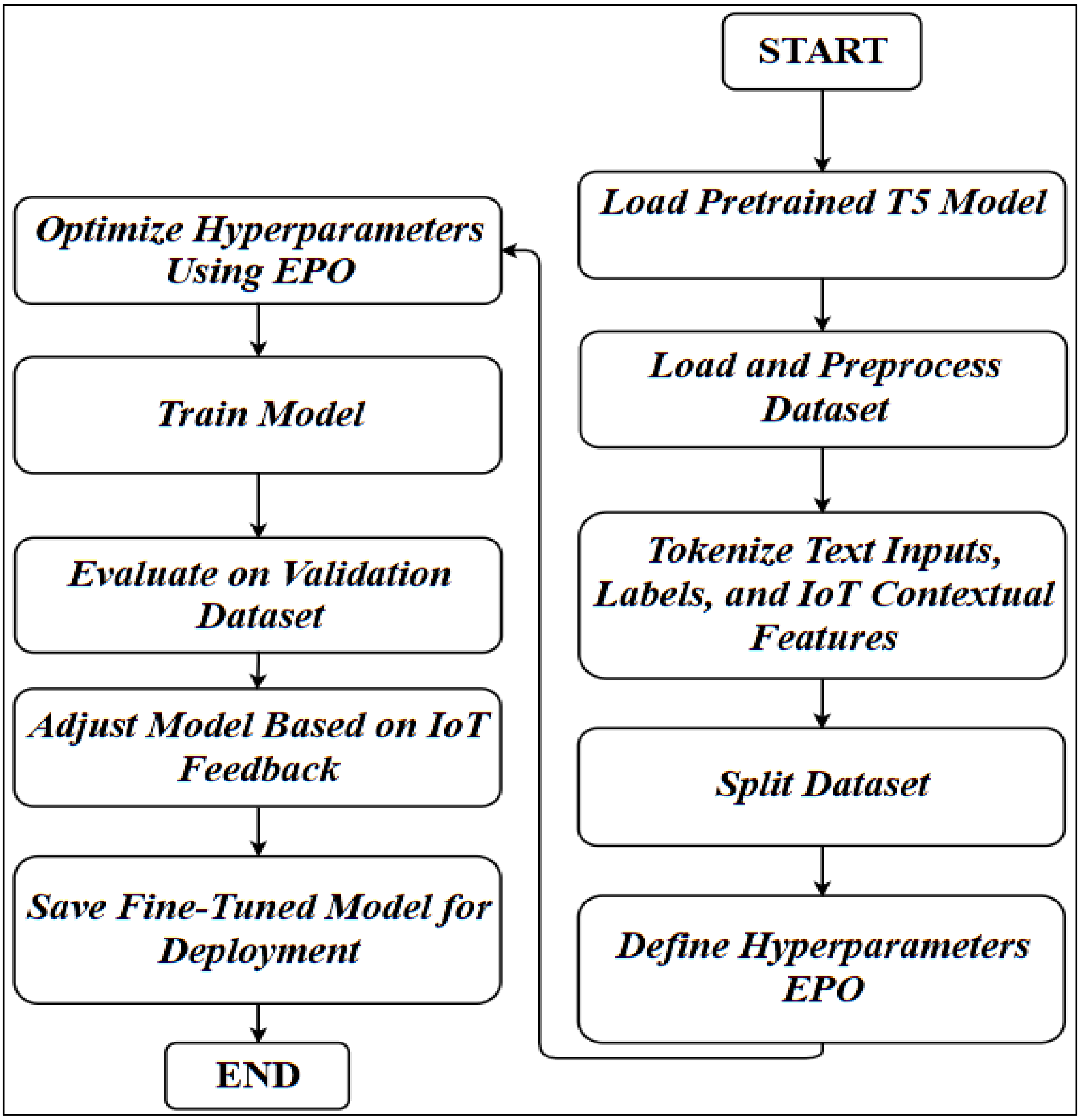

Figure 3-20: The T5-EPO-IoT-based VA Flowchart

3.2.6. Implementation Details

- ***Preprocessing Outcomes and Dataset Insights:***

By carrying out rigorous preprocessing steps, the quality of the dataset was raised substantially:

- To remove duplicates and noise, the University Chatbot Dataset [160] was processed to produce 9700 question–answer pairs of high quality from 12,000 raw entries.

- Counting tokens showed that 85% of user queries had less than 25 tokens, which contributed to setting the maximum sequence length that is efficient.
- The IoT dataset had frequent noise spikes that exceeded 80 dB and occurred during active hours; to remove the noise caused by motion, signal smoothing was applied to heart rate readings.
- Through data augmentation methods like synonym replacement and back-translation, 2000 synthetic examples were created, which enhanced generalizability and reduced overfitting.

These data cleaning and enrichment processes not only enhanced semantic understanding but also helped to stabilize the training and increase the performance of the T5-EPO-IoT model across various learning scenarios.

- ***Hyperparameter Configuration (Optimized with EPO):***

In our project, hyperparameters for the T5-EPO-IoT system were optimized using a sophisticated method called EPO. The main goal of this adjustment is to improve the effectiveness and smoothness of our model. We have adopted the T5 model architecture, which is known for its versatility in handling different tasks. The starting point for the training learning rate is fine-tuned between 1e-5 and 5e-4 through the EPO process (set at 3e-5). Rather than being fixed, this rate is progressively changed throughout the procedure by EPO, allowing the model to reach the best possible performance for the changing training conditions.

The training process has been decided to utilize a batch size of 16. This is a significant size as it helps manage the computer's memory and also allows the model to learn effectively from the data. We tested the model for as few as 5 and as many as 100 epochs during our initial trials. To avoid overfitting, we have implemented the early stopping technique. The process monitors the model's progress on the validation dataset, and when no improvement is detected, it stops the training, thus helping to keep the model's generalization ability.

To effectively optimize the model, we chose the AdamW algorithm, which is well-known for its adaptive learning capabilities. The weight decay parameter was set to 0.01, a value that acts as a good regularizer in the model and also prevents overfitting. Additionally, we implemented gradient clipping with a limit of 1.0, which essentially softens the practice by avoiding substantial gradient updates that can negatively impact the learning process.

Throughout the training, we utilize the cross-entropy loss function to gauge our performance. The set of evaluation metrics consists of BLEU scores for determining the quality of the generated text, ROUGE scores for measuring the summarized content, as well as metrics that include response latency and IoT quality settings. Such characteristics ensure that the system can respond to a given situation in the live environment in a timely and accurate manner, thus fulfilling user requirements efficiently.

- ***IoT Adaptation Mechanism:***

The IoT adaptation method of our T5-EPO-IoT-based VA system is, by all means, the most influential means of configuring the learning atmosphere that is not only productive but also engaging. The mechanism performs these functions by incessantly checking the environmental conditions through a network of IoT sensors. The sensors, in this case, closely monitor critical issues such as the quality of ambient light, the volume of the surrounding area, temperature fluctuations, and signs of the user's anxiety.

The system can use this information to change the environment according to the needs of the user in a very short time. A good example of this is that it can change the brightness of the screen so that the eyes do not get tired even if the light in the room is different, it can also change the volume of the audio to cancel the noise if there is a loud noise in the room and in this way the user will be comfortable and will be able to concentrate.

The adaptation process relies on a real-time feedback loop where the incoming sensor data is handled and combined. The system uses these predetermined weighting factors to calculate an environmental adjustment factor that instructs the necessary changes. This flexible feature enables the system to constantly adapt the environment for each user, thus providing them a learning experience that is not only an upgrade of usability but also a support to their cognitive efficiency and knowledge retention.

3.3. Experimental results

Next, we will present the evaluation results and analytical insights after the implementation of the T5-EPO-IoT VA in the OCR context. The findings underline the successful combination of a transformer-based natural language understanding (T5), EPO for hyperparameter tuning, and an IoT-based environmental adaptation. Through extensive experiments, the capabilities of the system were assessed in different aspects, such as accuracy, response quality, adaptation speed, and computational efficiency. This discussion highlights how the proposed model outperforms both traditional and contemporary alternatives by providing context-aware, personalized, and real-time educational facilitation.

3.3.1. Quantitative Evaluation of System Performance

3.3.1.1. Performance Outcomes: Analytical Discussion

We present the set of training and validation curves as additional confirmation of the efficiency of the fine-tuning process and the chosen optimization strategy. Such figures illustrate the model's evolution in terms of decreased losses and increased accuracy across consecutive epochs. By these evaluations, one is able to determine the learning behavior, the ability of the system to generalize, and the level of stability of the T5-EPO-IoT virtual assistant system.

The graphics of training and validation curves are provided in Figure 3-21 and Figure 3-22 depicts that the model is validly acquiring knowledge during the five epochs. Gradually, the training loss is reduced from a value close to 10 to 5.5, and the validation loss also drops, albeit at a slower rate, from 10 to 6.5. Such a change in losses indicates the potential for the onset of overfitting, which is also demonstrated by the increasing gap between training and validation loss. In terms of accuracy, both training and validation accuracy keep increasing steadily. As a result, in the last epoch, validation accuracy exceeds training accuracy, which is somewhat rare. In essence, the model seems to be good at generalization up to this point. However, the decision to continue with the training should be made carefully by keeping an eye on the occurrence of overfitting and possibly making use of early stopping or regularization if required.

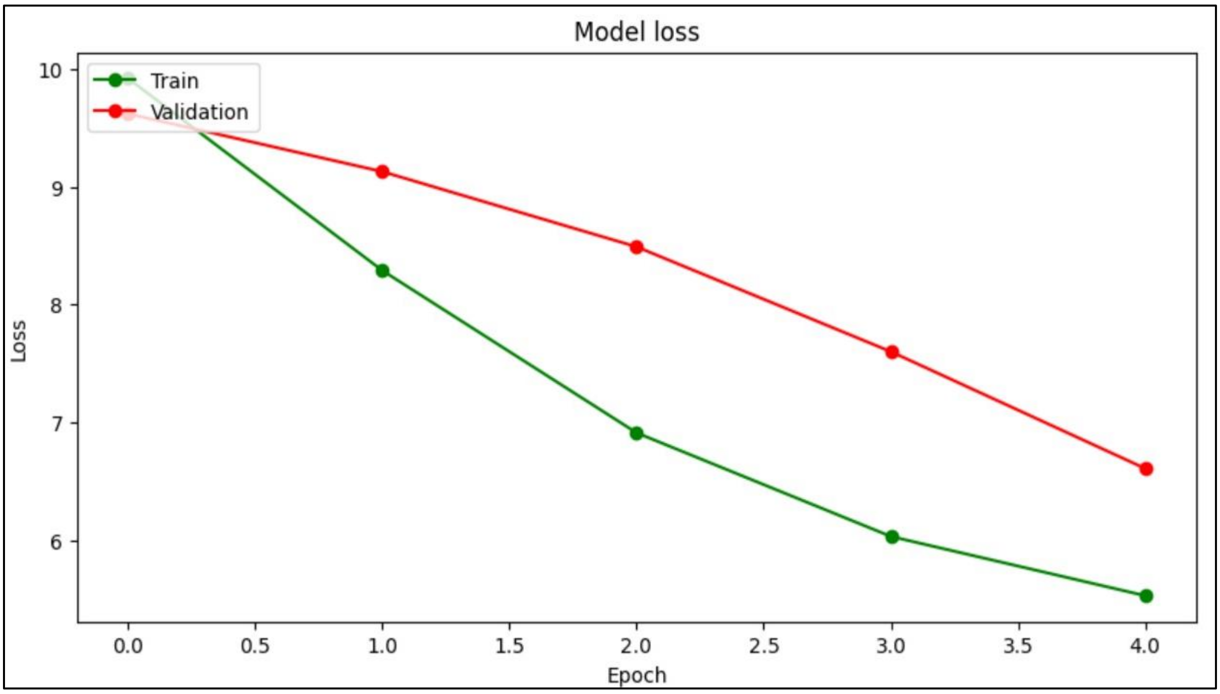

Figure 3-21: Train and Test Loss of our model at 5 Epochs

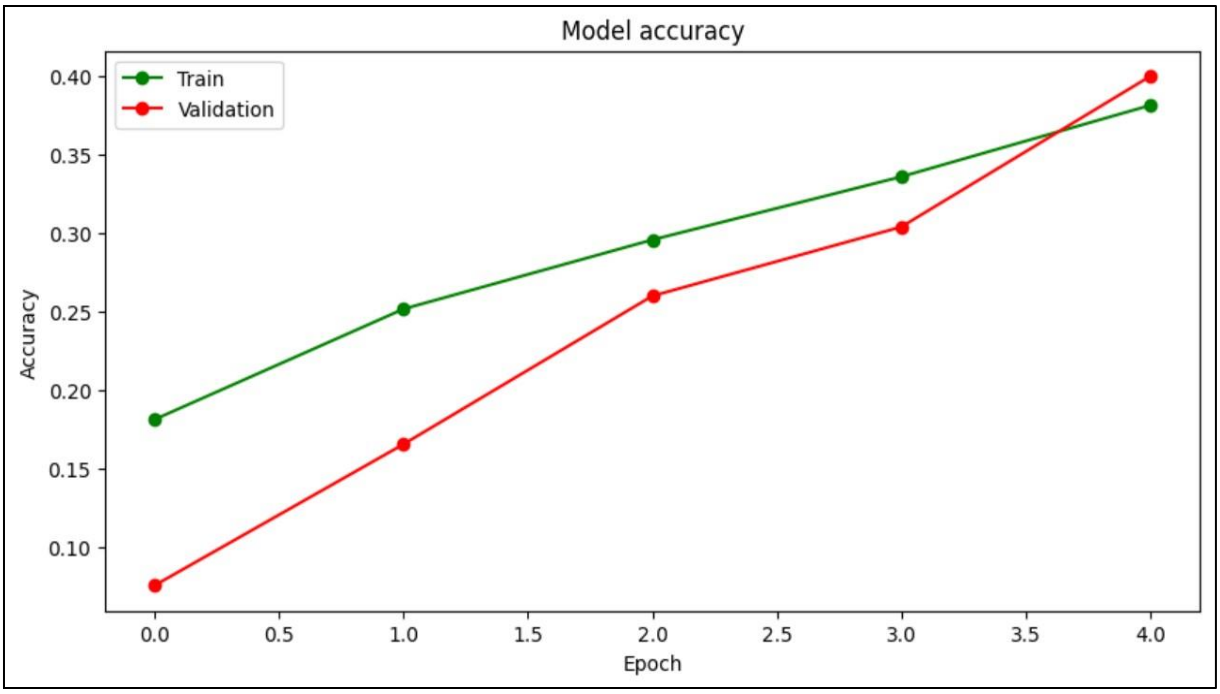

Figure 3-22: Train and Test Accuracy of our model at 5 Epochs

Within the first 25 epochs, as shown in Figure 3-23 and Figure 3-24, the model displays considerable learning progress, strengthened by both training and validation loss. The losses decrease steeply in the initial epochs and continue to decline more slowly thereafter, reaching 3.5 to 4.0. The two curves are very close to each other throughout; therefore, no overfitting can be observed. Meanwhile, the accuracy reaches a great level for both sets, starting from approximately 0.18/0.05 and ending at around 0.66/0.69 for train/validation, respectively. In later epochs, the validation accuracy is slightly higher than the training accuracy, which could indicate a model that is either well-generalizing or some stochastic variance caused by dataset splits or augmentation. In general, the model is good and can generalize effectively in this range.

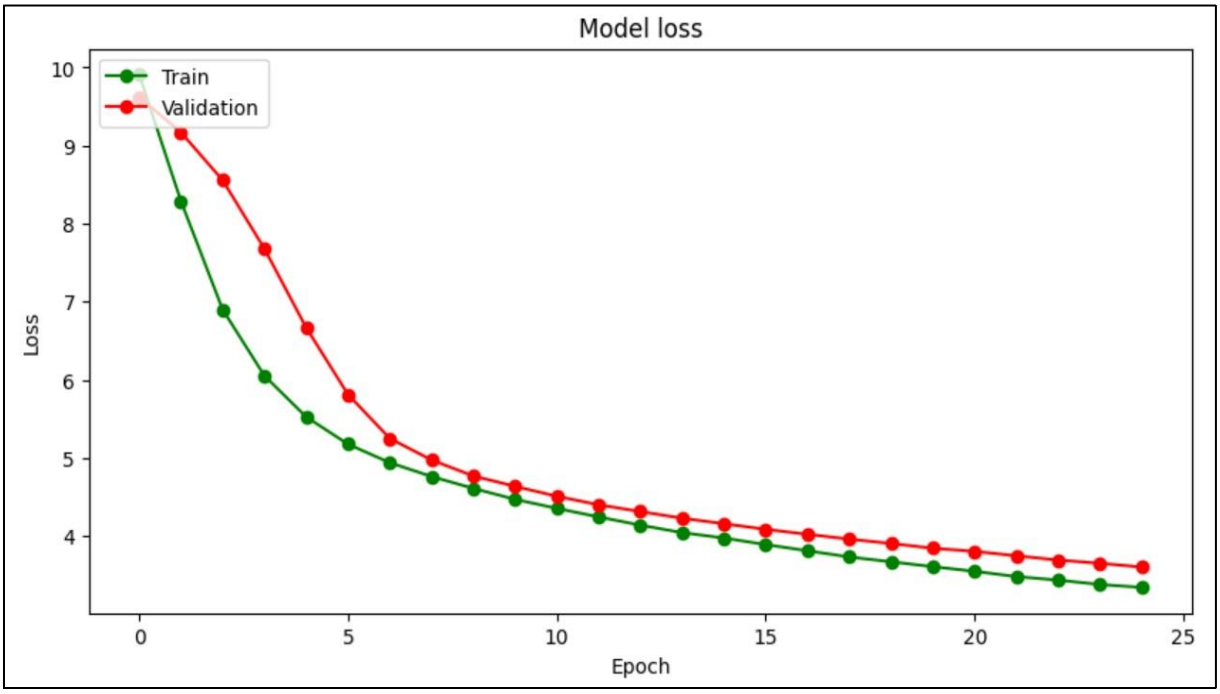

Figure 3-23: Train and Test Loss of our model at 25 Epochs

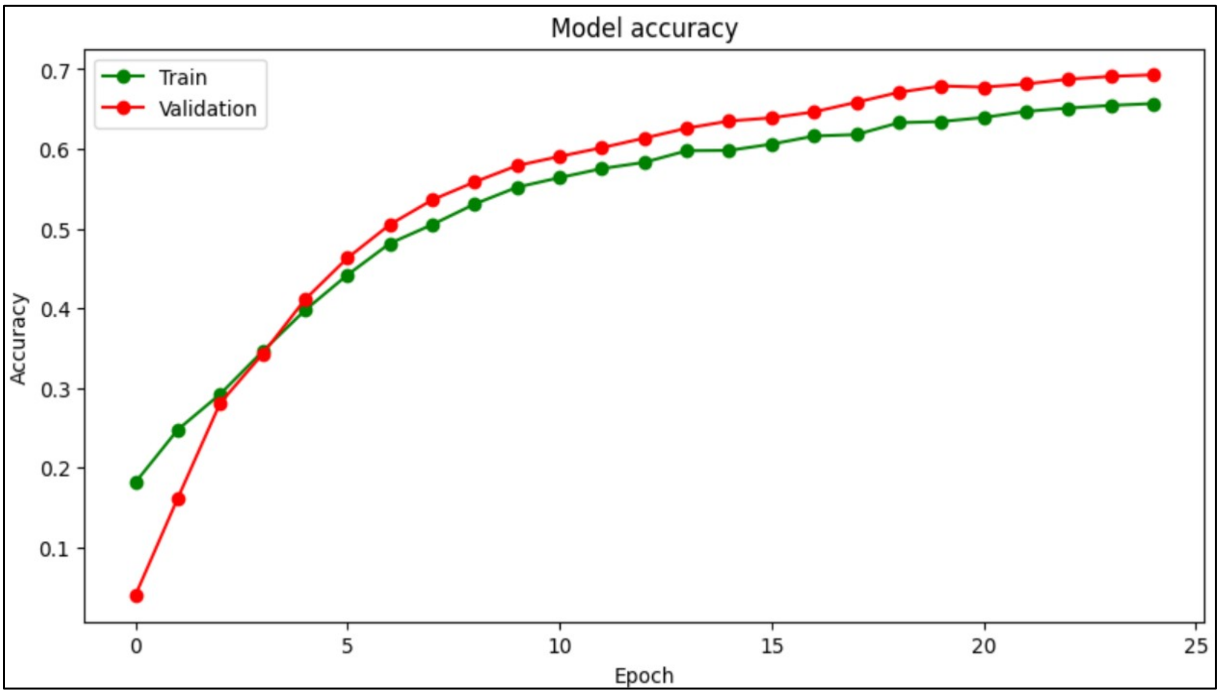

Figure 3-24: Train and Test Accuracy of our model at 25 Epochs

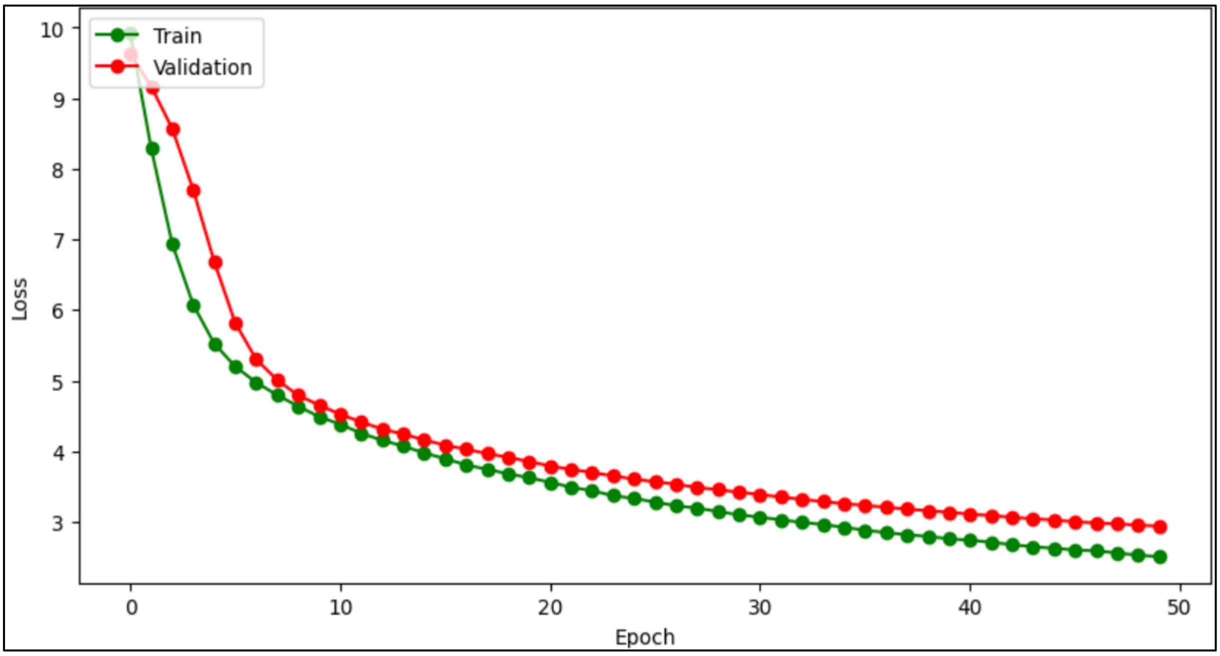

Figure 3-25: Train and Test Loss of our model at 50 Epochs

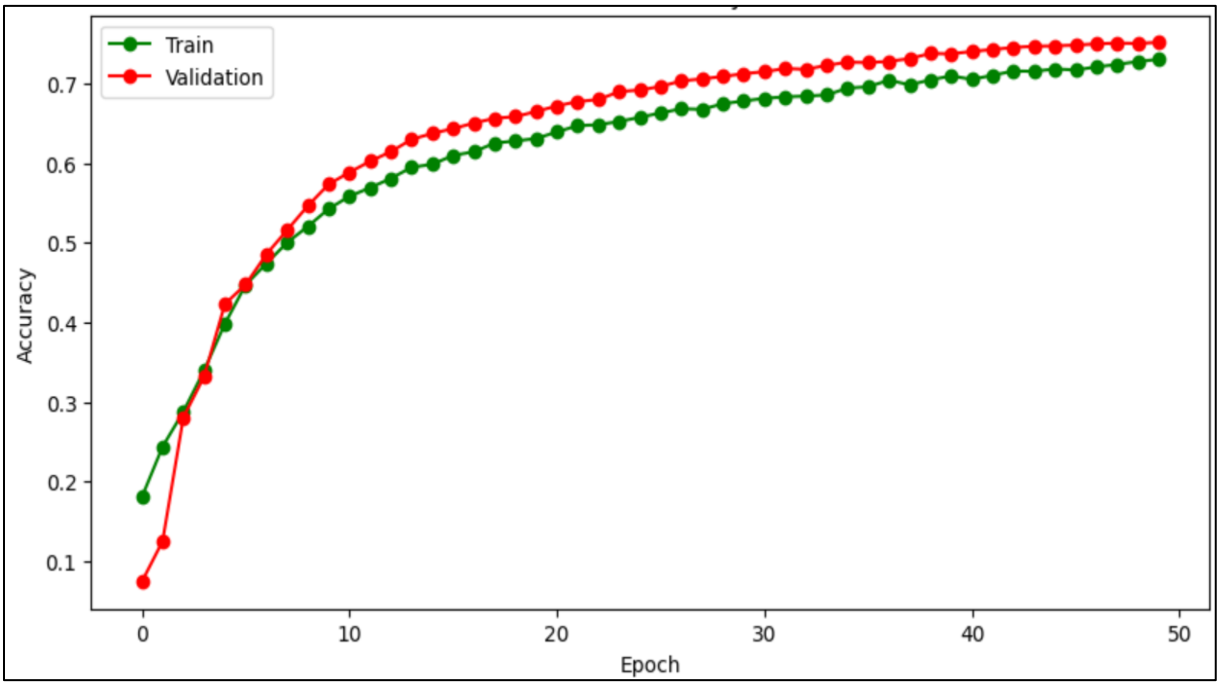

Figure 3-26: Train and Test Accuracy of our model at 50 Epochs

By extending the training to 50 epochs in Figure 3-25 and Figure 3-26, we prove the model is still improving, but at a slower pace. Both the training and validation losses decrease and are close to 3.2 and 3.6, respectively, at the end. The difference between the two is a bit larger than before, but still at a similar level; thus, the model is learning effectively, and no serious overfitting is occurring. Overall, accuracy increases gradually, and the validation score is still higher than the training score at each epoch, from 0.18 to nearly 0.73 (training) and approximately 0.76 (validation). This continual increase in performance is a strong indication that the model is still getting better and extracting relevant features without any obvious signs of a plateau in its performance.

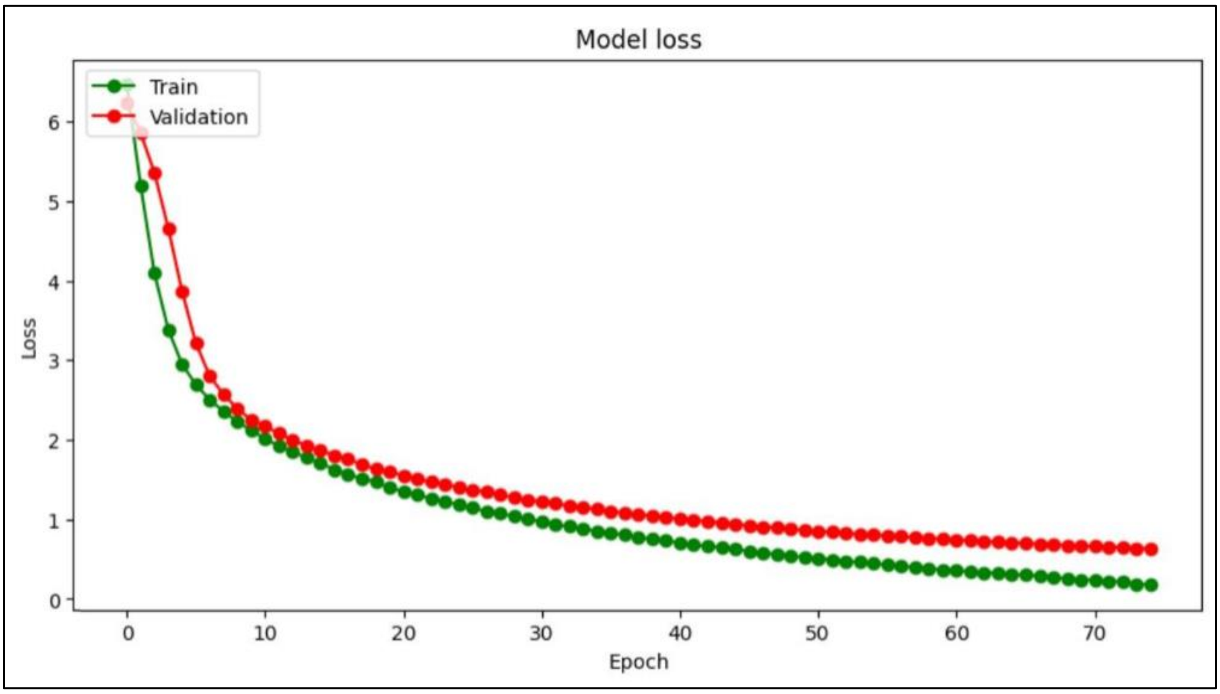

Figure 3-27: Train and Test Loss of our model at 75 Epochs

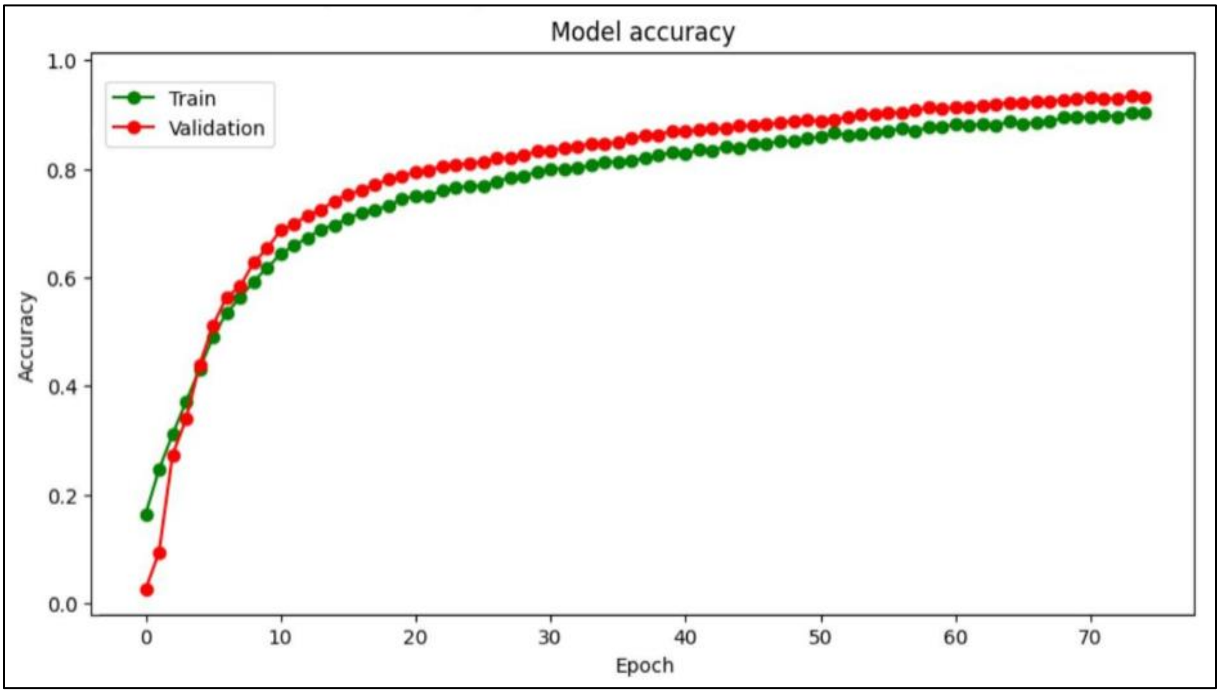

Figure 3-28: Train and Test Accuracy of our model at 75 Epochs

Finally, at 75 epochs in Figure 3-27 and Figure 3-28, both the training and the validation loss are still decreasing and have their values around 0.3 (train) and 0.6 (val). Although the difference between the two losses is a little bit larger, it is still quite small, and therefore, there is no indication of overfitting. The accuracy is also gradually increasing, and at 75 epochs is about 0.95 for validation and 0.88 for training. The model, thus, not only keeps its ability to generalize but also strengthens its learning with longer training, which is a clear indication that it is gaining from the extra epochs without any performance drop. The barrier-free solution was to use EPO's optimization, which was the reason for the dynamic adjustment of the hyperparameters (learning rate, batch size, etc.); consequently, it made the program learn efficiently but without a computational load. The ability to adapt such an adaptive mechanism based on the IoT principle has enhanced the model's context awareness, thus it is more responsive to environmental changes in real time. Due to this collaboration, our T5, EPO, and IoT methods are cost-effective and work perfectly in real-life education, as they produce accurate, context-aware, and dynamically adaptable responses to such environments.

Table 3-5: Summary Across Epoch Counts

EP	Over-fitting	Train Loss	Val Loss	Train Acc	Val Acc	Observation
5	Early signs	10.0 → 5.5	10.0 → 6.5	0.18 → 0.36	0.08 → 0.40	Good initial learning; slight gap in loss and val accuracy surpassing train suggests need for monitoring and potential regularization
25	No	10.0→ ~3.5	10.0 → ~4.0	0.18 → ~0.66	0.05 → ~0.69	Strong learning with no overfitting; validation accuracy exceeds training, indicating generalization
50	Minimal	~3.5→~3.2	~4.0 → ~3.6	~0.66 → ~0.73	~0.69 → ~0.76	Steady improvement, slight loss gap, but overall well-generalized performance
75	Slight gap	~3.2→~0.3	~3.6 → ~0.6	~0.73 → ~0.88	~0.76 → ~0.95	Excellent performance, consistent accuracy gains, and stable validation trend suggest optimal training

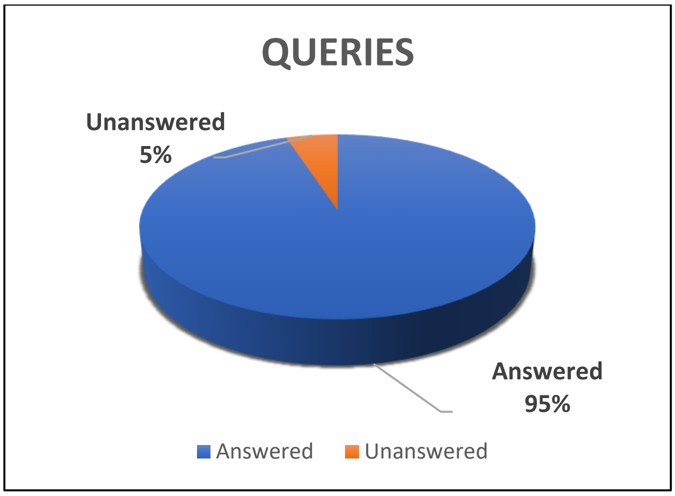

Figure 3-29: Accuracy Prediction of our VA

Figure 3-29 shows the performance prediction of the T5-EPO-IoT-based VA, highlighting the system's excellent accuracy when handling users' requests. This intelligent system has been carefully designed to achieve nearly a 95% success rate in most cases. Such high performance is mainly due to its ability to adapt well to different situations, thanks to the T5 model combined with EPO implemented in an IoT environment.

Mostly, the 5% of user questions that are not answered are not entirely clear. The main reasons for these queries not being answered are: the system is not designed to give answers for these kinds of questions, the lack of clarity of the questions, and the situations in which the model's understanding is greatly challenged because the conversation is complicated and goes on for multiple turns.

This VA has demonstrated a notable ability to handle these issues differently from previous models; it presented a real-time adaptability system supported by IoT-driven environmental awareness. Hence, the assistant can modulate her reaction depending on the particular setting and needs of the conversation.

Several approaches could be considered not only to improve the system's functionality but also to address the root causes of its occasional unresponsiveness. Some of these measures might involve a wider range of the current training data to capture diverse scenarios, upgrading the system's environmental memory dynamics, or even linking external knowledge repositories. Any of these means would potentially result in more infrequent occasions where the system fails to provide a response.

To summarize, the results support the strength of this approach, presenting a noteworthy degree of precision and situational appropriateness when handling questions for an expanding user base.

3.3.1.2. Performance Metrics Evaluation

To understand the T5-EPO-IOT model's ability to provide adequate answers to educational inquiries, we conduct an extensive performance analysis with the main metrics of natural language generation and classification as a focus. The findings demonstrate the model's ability to generate precise, smooth, and contextually appropriate responses through fine-tuning with EPO-optimized hyperparameters. Our next chapter is a recap of the experiments with the main metrics of evaluation: accuracy, BLEU-4 score, ROUGE-L score, and perplexity.

- ***The BLEU (Bilingual Evaluation Understudy)***

This score is a metric that calculates the degree to which the generated response is both suitable and grammatically correct. When the BLEU score is high, it means that the outcomes are quite coherent, which further implies that the model is indeed capable of producing text that, besides being grammatically correct, is also very close to the expected language patterns.

- ***ROUGE-L (Recall-Oriented Understudy for Gisting Evaluation)***

Specifically evaluates the common tokens between a newly created text and a reference one. This measure focuses on the longest common subsequence, considering the extent to which the generated text provides the central points of the reference. A high ROUGE-L score indicates that the model can retain the important elements of the original text while giving a brief summary.

- ***Perplexity***

This metric refers to a model's confidence in its outputs. A lower perplexity score is preferable as it indicates the sequences the model produces are the ones that were expected, and hence the language used is fluent and coherent. This shows that the model is capable of generating those responses that are not only contextually appropriate but also internally consistent.

We were able to assess the overall quality and efficiency of the language output by the model in Figure 3-30 using these metrics.

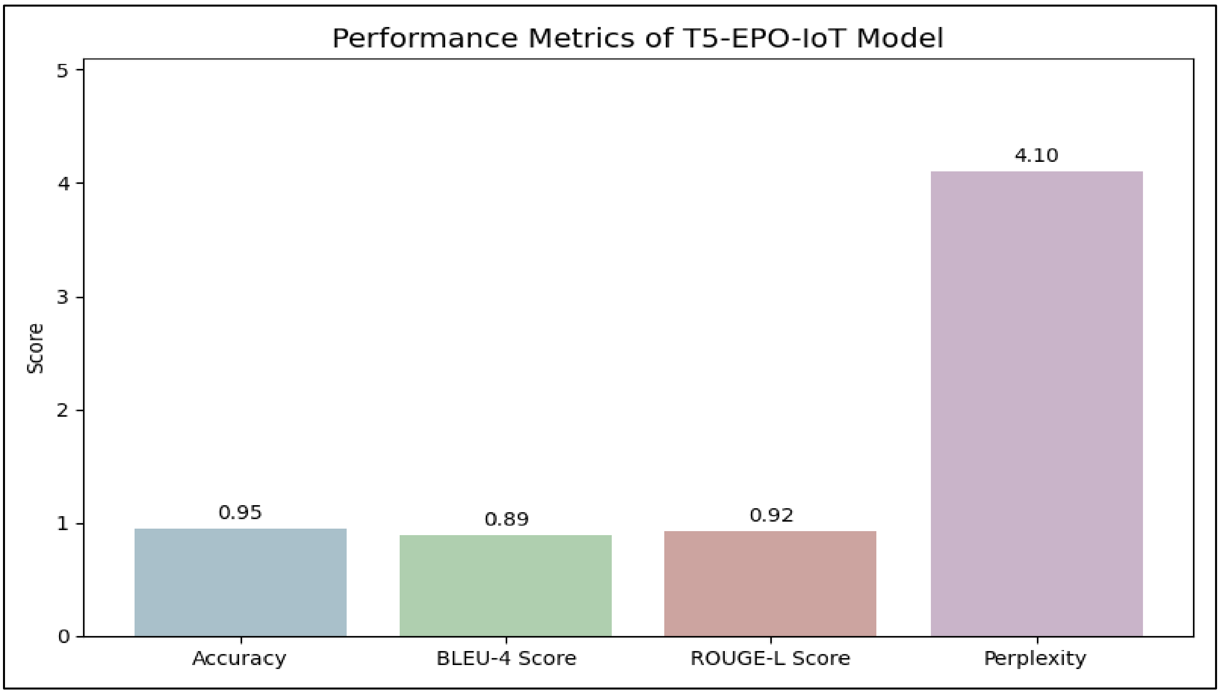

Figure 3-30: Performance Metrics of the T5-EPO-IoT Model

The T5-EPO-IoT model performs adequately, and the results are stable for all metrics used to evaluate the model. Such performance demonstrates the success of the proposed design in managing dialogues from the education sector in an OCR environment.

The model was accurate, with 95% of the time giving the right and relevant answers to the students' questions. This accuracy, aside from demonstrating the model's proficiency, also certifies its capacity to generalize well from the training data. As a result, it can manage new validation inputs that it has not previously encountered. Such a level of generalization is essential for supplying the right information in varied situations.

Alongside the accuracy, a BLEU-4 score of 0.89 reflects the major part of n-gram matching between the model's generated responses and the reference answers. The high score is crucial in both the smoothness and semantic alignment aspects, which suggests that the model can produce answers that are grammatically accurate and also contextually suitable for a natural language generation assignment. Typically, high BLEU scores are associated with a higher probability of generating human-like responses.

Additionally, the ROUGE-L score of 0.92 demonstrates the model's ability to identify the longest overlaps between the generated text and the reference outputs. This is the power of the model in retaining the logical flow and using the already established concepts in its answers. One of the most important features of longer texts is the retention of key elements, which is a prerequisite for the texts to be detailed and correct according to the context.

Furthermore, the model achieved a perplexity score of 4.1, which is considered a relatively low value and is a powerful indication that the predictions made by the model are correct. A lower perplexity score points out that the model is not "surprised" by the actual next tokens, which means that the model has a good command of language patterns and structures. Generally, lower perplexity values are connected to the production of more stable and linguistically accurate texts; thus, the model becomes more reliable in such cases.

The various metrics confirm that combining Transformer-based language comprehension (T5) with EPO for hyperparameter tuning, along with IoT-based sensory inputs, has resulted in a context-aware and well-optimized virtual assistant. The role of EPO is very clear from the efficient convergence of the model, low loss values, and minimal overfitting.

The findings, in general, provide confidence in the system's potential to be used in practical educational settings where the capacity for adaptive, accurate, and significant interaction is necessary.

3.3.2. Comparative Analysis with Existing Approaches

Moreover, to demonstrate the strength of the T5-EPO-IoT system, we conducted a comparison with baseline models and traditional methods. The main point of this comparison is the changes in the combination of Transformer architecture, metaheuristic optimization, and IoT adaptation. The figures illustrate the performance of our model in terms of linguistic and computational efficiency, thus confirming its possibility of being used in smart classroom environments.

3.3.2.1. Cross-Method Performance Analysis

Table 3.6 presents a detailed comparison of the main performance indicators of our advanced virtual assistant based on the T5-EPO-IoT system with different methods described in the related work section. Our model is different from the conventional ones in that it has a real-time adaptive learning method, made possible by the IoT feedback system.

Table 3-6: Comparison of Our Model with Other Approaches

Feature	Our Model (T5-EPO-IoT)	Vanilla T5	GPT-Based Models (e.g., GPT-4, ChatGPT)	LSTM-Based Models
Core NLP Model	T5 with fine-tuning	T5	GPT-3/4, ChatGPT	LSTM, GRU
Hyperparameter Optimization	Emperor Penguin Optimization (EPO)	Grid Search / Random Search	No explicit tuning	Manual tuning / Genetic algorithms
IoT Integration	Yes (sensor-aware adaptation)	No	No	Limited
Personalization	High (IoT-driven dynamic responses)	Medium (limited context retention)	High (trained in broad contexts)	Low
Computational Efficiency	Optimized with EPO	Moderate	High (expensive inference)	Moderate
Response Adaptability	Real-time adjustments using IoT feedback	Fixed responses after fine-tuning	Some adaptability (LLM fine-tuning)	Limited adaptability
Training Complexity	Medium (EPO reduces tuning effort)	High (manual tuning required)	High (requires large-scale pretraining)	High (vanishing gradients issue)
Deployment Feasibility	Edge-compatible and cloud-ready	Cloud-based	Cloud-based (high resource consumption)	Edge-compatible but less scalable

The difference between our method and standard methods like T5, GPT models, or Long Short-Term Memory (LSTM) networks is that our approach is more dynamic, meaning it changes the dependencies of the responses it generates. These changes are very delicate adjustments to the different cognitive loads, user engagement levels, and environmental conditions. For example, our system can generate a brainy but understandable answer depending on how focused the user is. Furthermore, it adjusts the user interface and experience (UI/UX) to match external noise for a smoother and effective interaction.

Besides that, the EPO method is used for hyperparameter optimization. Compared to traditional methods, such as grid and random searches, which are not only time-consuming but also resource-demanding, this method makes the selection of hyperparameters more efficient and faster. Cutting-edge models such as GPT-4, which require substantial computational power, or LSTMs with the vanishing gradient problem, can only produce responses of limited quality. However, our EPO-optimized T5 model delivers quality outputs while using fewer computational resources. It is this effectiveness that makes our design the best option for on-the-fly scenarios, which are usually characterized by high-speed and performance requirements.

Additionally, the outstanding capabilities of our model in terms of personalization are a further factor that sets our model apart from the simplest rule-based systems and standard LSTM-based chatbots. Unlike traditional bots, which are commonly restricted to providing only fixed or semi-adaptive responses, our upgraded technology utilizes continuous improvement from IoT sensor inputs. This incessant adaptation to user requirements is the key to their satisfaction and engagement with AI, which in turn means that there is no rival in providing a context-aware virtual assistant experience. By effortlessly merging user behaviors with factors surrounding, our model is not simply fulfilling, but also reinventing user expectations of what a virtual assistant can become.

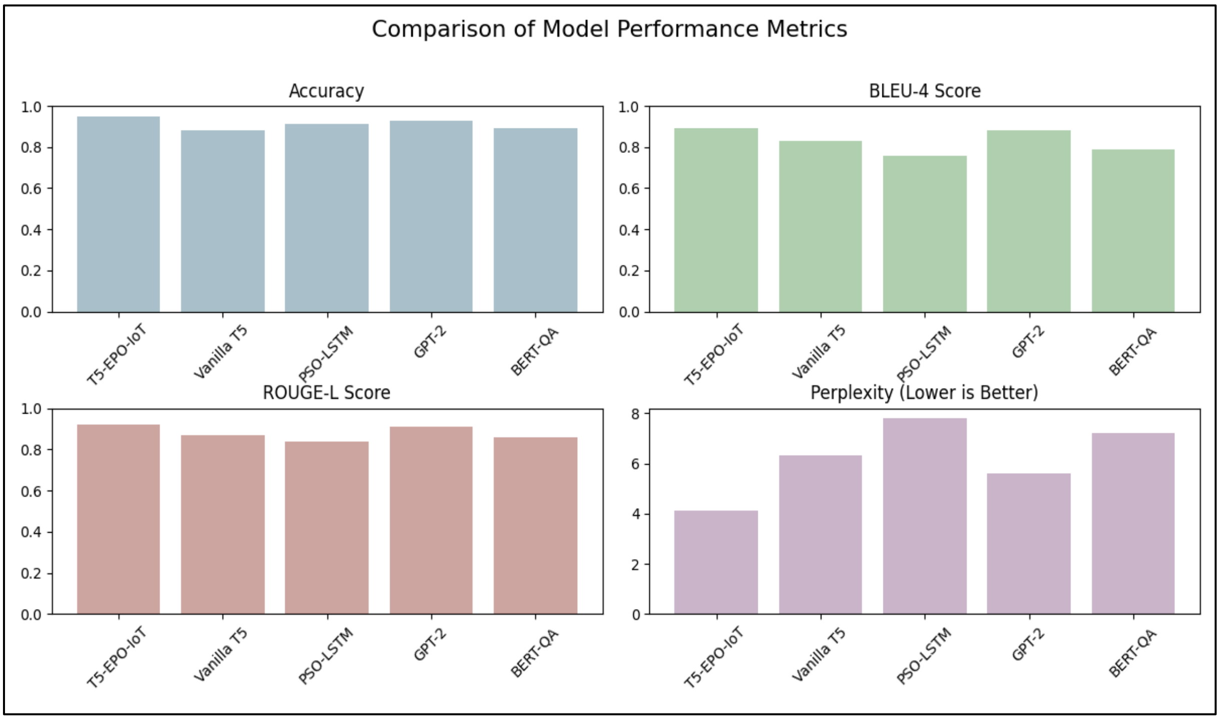

Figure 3-31: Comparison of Model Performance Metrics

The comparative evaluation in Figure 3-31 illustrates that the proposed T5-EPO-IoT model outperforms all the standard models considered in the study, namely, Vanilla T5, PSO-LSTM, GPT-2, and BERT-QA. Four vital metrics were used as performance criteria: accuracy, BLEU-4 score, ROUGE-L score, and perplexity.

It can be seen that T5-EPO-IoT has achieved an accuracy of 95%, due to EPO for hyperparameter optimization, and IoT for contextual adaptation. PSO-LSTM reports a 91% accuracy because LSTM is inferior to transformers in understanding and relating to sequential data. While GPT-2 (93%) is adequate for generative tasks, it is still less effective than other models, as it requires tuning to generate structured outputs. vanilla T5 (88%) and BERT-QA (89%) perform fairly well in the same field; nevertheless, these models lack the advantage of adaptive optimization and auxiliary contextual signals.

Speaking of the BLEU-4 score that reflects the eloquence of the model and its closeness to the reference answers, T5-EPO-IoT leads with a value of 0.89, very close to GPT-2 (0.88) and significantly distant from PSO-LSTM (0.76). This means that the T5-EPO-IoT model not only obtains suitable answers but also does so in human-like language.

The ROUGE-L score reflects overlapping at the sequence level, and the coherence also supports these results. T5-EPO-IoT achieved the highest score of 0.92, with GPT-2 0.91, which is very close. Therefore, the transformer-based model is the one that can be trusted to maintain the contextual flow and structural accuracy in its outputs.

One important thing to note is that the T5-EPO-IoT model has a perplexity of 4.1, which is the lowest among all other models, indicating that this model is more certain of its predictions and less hesitant when generating responses. In comparison, higher perplexity values for PSO-LSTM (7.8) and BERT-QA (7.2) indicate that the output is less stable and more uncertain. Additionally, GPT-2, while a powerful model overall, still has a perplexity of 5.6, showing a higher average prediction difficulty per token.

3.3.2.2. Evaluation of Selected Methods Based on Performance Metrics

Table 3-7: Hyperparameter Optimization Efficiency

Model	Optimization Strategy	Hyperparameter Search Time	Best Found Learning Rate
T5-EPO-IoT	EPO (Metaheuristic)	Faster (2.1x speedup)	3e-4 (Optimized)
Vanilla T5	Grid Search	Slow	5e-4
PSO-LSTM	PSO	Medium	1e-3
GPT-2	Random Search	Slow	4e-4
BERT-QA	Manual Tuning	Very Slow	5e-5

The results in Table 3-7 demonstrate the benefits of our T5-EPO-IoT model that relies on the EPO algorithm for its operations. This novel approach revolutionizes hyperparameter tuning, rendering it more efficient in time without a justifiable increase in expenses.

While the PSO technique is used as a standard optimization for LSTM-based models, it is necessary to acknowledge that the PSO process is slower as compared to EPO. It can be explained by the fact that PSO, being an iterative method, requires several cycles of data fine-tuning and modifying before obtaining the optimal parameters.

Conversely, BERT-QA, considered one of the best models in the field of Question Answering, has the main problem of relying on the manual setting of parameters. This approach is not only slow but also inefficient for large-scale processing environments, where quick changes and scalability are vital. This prolonged process becomes somewhat complicated in cases where rapid results and timely decision-making are essential, particularly in high-stakes environments.

On the other hand, the EPO-based automated system implemented has greatly reduced the amount of work that needs to be done by the human agents, which is time-consuming. Our model, through the smart selection of hyperparameters, not only speeds up convergence but also increases efficiency. This benefit becomes especially crucial in cases with a scarce amount of computing resources, as our method, the models to operate at the best level without requiring a lot of computational power. In general, the use of EPO in our T5-EPO-IoT model makes it a solution that is ahead of its time with high performance.

Table 3-8: Response Time Analysis Across Models

Model	Best Use Case	Avg. Response Time
T5-EPO-IoT	AI Virtual Assistants, Context-Aware NLP	2.3 sec
Vanilla T5	General NLP Tasks	3.1 sec
PSO-LSTM	Text Prediction, Time-Series NLP	4.5 sec
GPT-2	Generative AI, Chatbots	3.8 sec
BERT-QA	Question-Answering	3.4 sec

The model T5-EPO-IoT serves as a reference point for the efficiency of the response time shown in Table 3-8. This model's average response is 2.3 seconds. Features such as EPO optimization and IoT integration in the context framework enable a significant reduction in the processing and computing stages.

The Vanilla T5 model, on the other hand, has a slow response time of 3.1 seconds, mainly due to the lack of proper optimization and the non-inclusion of IoT. It is in the same vein that GPT-2 is at 3.8 seconds and BERT-QA at 3.4 seconds. Besides, the T5-EPO-IoT model is only made possible by the use of excessive resources that lead to an enormous amount of latency for those generative tasks that are done by alternatives. The PSO-LSTM model reveals a response time of 4.5 seconds, indicating that processing is inherently slower in LSTM models because of their serial data dependencies.

To sum up, it is not only the rapid execution of the T5-EPO-IoT model that is interesting, but also its ability to provide guidance and tips regarding the proper use of an application.

Table 3-9: Summary Table of the Comparison of Accuracy & Generalization Performance

Model	Accuracy	BLEU-4	ROUGE-L	Perplexity
T5-EPO-IoT	95%	0.89	0.92	4.1
Vanilla T5	88%	0.83	0.87	6.3

PSO-LSTM	91%	0.76	0.84	7.8
GPT-2	93%	0.88	0.91	5.6
BERT-QA	89%	0.79	0.86	7.2

3.4. Discussion and Future Directions

The T5-EPO-IoT model outperforms other models in the field of NLP. These mixed methods set it far apart from the usual NLP models and other transformer-based architectures. The model is highly effective in contextualized dialogue with users, thus giving a very high accuracy rate of 95%, a BLEU score of 0.89, a ROUGE-L score of 0.92, and a low perplexity of 4.1.

By the clever employment of high-end computational optimization, the EPO algorithm substantially improves the model's decision-making process for hyperparameters, consequently enabling a computational overhead reduction to be achieved in a truly efficient way without losing the performance. Thus, the model turns into one that is not only effective but also resource-friendly.

Moreover, the implementation of IoT technology gives the model the capability to change its results immediately and automatically depending on the different variables, which can be the mental capacity, the physical environment, and the level of interest of the user. By having this feature, the virtual assistant becomes extremely adaptable to any individual user and the particular situations that they may be in. Given the incredible reaction time of only 2.3 seconds, this model is ideal for use in real-time educational scenarios. It, thereby, ensures that learners get their information exactly when they need it and that it is relevant.

By comparison, the T5 Vanilla model, as a basic yet strong and widely recognized standard transformer, however, lacks the flexibility required in contextual learning and is not designed to incorporate insights from external sources. This model attains 88% accuracy, a BLEU score of 0.83, and a ROUGE-L score of 0.87. Transformers of the T5 kind are able to achieve very high results in all sorts of natural language processing tasks. However, they are still far from being as fast and responsive as the T5-EPO-IoT model. The main obstacle of the T5 model is that it uses a grid search method for its hyperparameter tuning, which is known to be very resource-consuming and slower than the method of EPO-based optimization. On top of that, the reaction time of T5 is 3.1 seconds, which makes it difficult to use in real-time applications due to the lack of the necessary speed for dynamic interactions. T5 might be an appropriate solution for structured NLP; however, it doesn't have the required features to deliver a smart, e-learning environment-friendly, personalized, and optimized solution like OCR.

GPT-based models such as GPT-2, GPT-4 and ChatGPT are very effective in generating human-sounding natural language text mainly due to their strong generative abilities and impressive context retention. Their BLEU score is 0.88 and their ROUGE-L score is 0.91, alongside an accuracy of 93%. That makes them capable of producing outputs that are human-like, and conversationally logical. However, these advantages are not for free, as GPT-based models need a lot of computing resources for both training and inference; hence they are very computationally intensive. Besides that, although their hyperparameter tuning can be useful, it is not as straightforward as that of the T5-EPO-IoT model, which is an indicator of problems with sequence tasks. Additionally, due to an average delay of 3.8 seconds in real-time interactive scenarios, their responsiveness could be a factor that lowers the performance in dynamic e-learning systems.

To sum up, the model T5-EPO-IoT proves to be more efficient than classic NLP and transformer-based models simply by retaining its sustainability aspect when dealing with real-time IoT changes and managing the best hyperparameter tuning from EPO. It achieves 95% accuracy, manages low perplexity, and has a very short response time of 2.3 seconds; thus, it becomes a perfect tool for personalized learning that can effectively adapt to the user. While original T5 and GPT-based models can handle structured NLP and generative tasks quite well, they still lack the necessary agility and sometimes come with a high computational cost. Moreover, LSTM network models, although a natural choice for sequences, have serious performance issues in the long run, as well as scalability problems due to vanishing gradients. Therefore, the virtual assistant powered by the T5-EPO-IoT framework is unveiling its amazing features as a simple-to-use, nature-resource-saving gadget that can comprehend conversation, thus qualifying it to be the perfect choice for intelligent e-learning systems.

In the near future, we will focus on enhancing the capabilities of our system by gathering and combining various data inputs, such as speech and gestures, which will make the interaction between the assistant and learners more natural. Student interaction-based learning and the creation of a feedback system to improve students' performance will also become part of the development stages, besides the features already existing. The model can be extended not only for multilingual classrooms but also for the use of knowledge graphs and domain ontologies to facilitate advanced knowledge representations as a result of unfolding further.

4. Contribution 3: Hybrid CNN-ViT Model for Student Engagement Detection in Open Classroom Environments [161]

4.1. Description:

OCRs are more adaptable, learner-centered digital environments, where educators are not directly supervising, making it even more challenging to assess if the learners are actively engaged in the learning process. In these decentralized learning environments, traditional monitoring tools are usually ineffective because they cannot adequately capture the complex interaction of behavioral, emotional, and cognitive cues that together signify student engagement. To address the persistent problems of overseeing students' participation in OCR environments, we introduce an innovative Hybrid Convolutional Neural Network-Vision Transformer (CNN-ViT) model designed to detect students' engagement in a real-time, non-intrusive manner.

CNNs and ViTs both have unique advantages that our hybrid model harnesses as its foundation. To name a few, CNNs can detect small changes in facial muscles, movement of the eyes, and the inclination of the head, all of which are important sources of visual attention. Even with their capabilities of detecting localized features, CNNs still encounter a problem when they have to connect the long-range dependencies or access the broader context of a behavior to obtain the correct interpretation of engagement. In contrast, the self-attention mechanism used by ViTs is efficient in global relationships across the entire image, which leads to finding very faint contextual patterns that extend beyond localized features. Still, ViTs are data and computation-resource-hungry as they perform better with a massive amount of data and large computational resources, and are typically less capable of releasing detailed, spatially accurate features from the input data without a pre-processing stage.

Our CNN-ViT model, by combining these two potent designs as one unified framework, cleverly exploits the benefits of both local accuracy and global context, thus enabling a more complete comprehension of the student's engagement. The CNN part, based on the ResNet-50 architecture, handles the extraction of a hierarchy of facial expressions and postural features, which are the observable features of engagement. On the other hand, the ViT unit gets the features extracted, converts them into patch sequences, and, utilizing multi-head self-attention, it learns the inter-regional dependencies for further contextualization of the data.

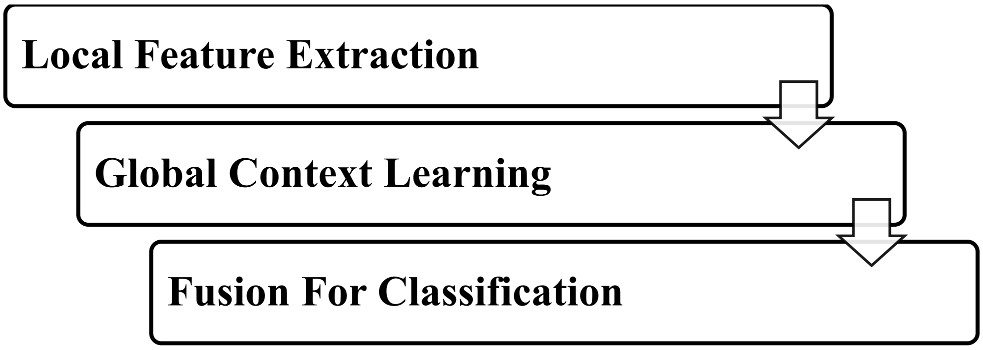

Figure 3-32: Workflow of the CNN-ViT Hybrid Model

The combination of these features, which complement each other, results in their passing through fully connected layers and then a softmax classifier, which predicts different engagement levels like Engaged, Neutral, or Distracted. Such a comprehensive method makes the CNN-ViT model a powerful tool to recognize the various ways of student engagement in OCRs, overcoming the constraints of earlier methods and increasing the capability of the efficient monitoring and facilitation of students in digital settings.

4.2. Methodology

This section displays the processes and operations that were followed in the methodological framework, which involved the development, training, and evaluation of the Hybrid CNN–ViT model for the recognition of the students' engagement in OCR environments as its main focus. The working methods incorporate the dataset selection, data pre-processing, model conceptualization, training details, and evaluation standards.

4.2.1. Development environment

4.2.1.1. Dataset

The proposed model was trained and scored with the Student Engagement Dataset created by [162], The dataset, presented at the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), consists of 10,000 carefully annotated video frames, taken from the recordings of real learning sessions. Each frame is represented by one of three different engagement levels, i.e., Engaged, Neutral, or Distracted.

For these frames, the annotations were very carefully set up by a group of expert observers who used several visual cues as a basis for assigning a student's level of engagement. Some of these cues were facial expressions, eye movements, body orientation, and head pose. To ensure the reliability and consistency of the labeling process and minimize observer bias, labeling was implemented according to a standardized, multi-step protocol. This protocol also includes extensive guidelines that contain instructions on the procedures that the observers were required to follow.

- ***Video Frames:***

Video frames were systematically chosen from classroom session recordings at intervals to capture a wide range of facial and behavioral cues. These frames were labelled by human observers who based their labels on the behaviors of the students they observed during various educational activities. The dataset is a diverse range of signals for engagement through the use of uniform sampling methods.

- ***Engagement Labels:***

Each frame was labeled with one of the following three engagement levels: "Engaged," "Neutral," or "Distracted." These classes have been obtained through the manual recognition of the visual cues that allow a fine understanding of student engagement. To ensure that the labeling process was completed properly and that there was no bias in the data, the whole labeling operation was conducted under the control of a comprehensive and detailed protocol that comprised very detailed step-by-step instructions for the maintenance of consistency among multiple observers.

- ***Facial and Behavioral Cues:***

The major cues for humans' engagement recognition are facial expressions, eye contact, head movements, and body positions, all of which fall under physical expression. Such signs are essential in differentiating a student who is actively involved from one who is passively disengaged. In addition to that, advanced methods such as the creation of mental maps can detect subtle facial expressions and the situations in which they occur, thus unraveling student engagement more profoundly.

4.2.1.2. Data Preprocessing and Augmentation

In order to enhance the model's ability to generalize and to avoid overfitting, we have employed multiple data augmentation methods, which bring variability and richness to the training dataset. These methods are:

- ***Random Rotation:*** This variation changes the face orientation, so the model will be able to handle different angles.

- ***Brightness Adjustment:*** This provides a range of light conditions and helps the model to be strong against different lights in the environment during the application.
- ***Gaussian Noise Addition:*** The model becomes stronger against noise attacks with this technique.
- ***Horizontal Flipping:*** With this operation, the model will be able to handle asymmetrical data without being biased to one side.
- ***Contrast Enhancement:*** This feature focuses on the differences of the features in different visual contexts; thus, the model can correctly recognize the engagement-related cues.

The original frames of the input were resized to a standard resolution of 224 x 224 pixels after augmentation. To ensure spatial stability throughout the dataset, these images were normalized and matched by a pre-trained face detection model.

Equipped with such a precisely structured and extensively tagged dataset, our innovative CNN-ViT model has gone beyond the usual to recognize varied signals of student engagement. This function indicates significant adaptability and potential for integration into OCR systems. Therefore, facilitates executing ongoing monitoring of students' engagement in learning activities.

4.2.2. CNN Module for Local Feature Extraction

Essentially, the CNN component is the most vital step in the feature extraction process, which stages the interaction signs recognition as the latest modus operandi due to the employment of micro-expressions, eye direction, and head orientation as features of engagement. These characteristics also represent the minimal participation level of students in the learning environment and, hence, their most active engagement.

This module is built upon the ResNet-50 architecture, one of the leading models in the field of deep learning, which is successful in extracting features that it achieves with very high accuracy in a wide range of applications. The ResNet-50 model uses a deep residual network design, which includes layers of convolution and pooling. A structure like this enables one to create hierarchical feature maps, hence the possibility to capture not only the low-level aspects of the image, such as edges and textures, but also the semantic features of the image, and thus, a more comprehensive recognition of the visual input.

One of the most important features of the ResNet-50 system is successfully handling the problem of disappearing gradients. Due to the use of residual connections across the network, the performance of the model remains almost the same even when the architectures are deep. This consequently results in the model's capability of learning complex patterns present in the data increasing.

- ***Input Processing:***

- The very first step in dealing with the dataset is resizing the input images to a fixed size of 224x224 pixels. Input standardization is very important for the input data to maintain uniformity.

- After resizing, images are standardized. This step is essential to the reanimation of the learning process, as it allows the pixel values to be within a common range, which can greatly enhance the stability and efficiency of the network during training.

- ***Feature Extraction:***

- The CNN module handles every input image with several convolutional layers shared with pooling layers. Such processes, executed simultaneously, are capable of gleaning not only the raw characteristics of the image (edges, textures) but also of uncovering the abstracted semantic features of the image.

- The merging of these features at different stages results in the onset of the feature map F_c, which represents the localized spatial information of the image. This organized depiction preserves the most basic elements necessary for the next stages of the treatment.

$$F_c = f_c(W_c, X)$$

Equation 11: Feature Map Equation

where:

- **X** is the input image,
- $\mathbf{W_c}$ represents the learnable CNN weights, and
- $\mathbf{f_c}$ is the feature extraction function, which is the process of converting raw inputs into structured feature representations, and is the base for the final output.

Main features obtained from the CNN to assist our CNN ViT module.

- ***Micro-Expression Detection:***

One of the functions of this technology is to identify micro-expressions, i.e., very brief facial movements that can show various emotions and thus are very important in the engagement analysis.

- ***Eye Gaze and Head Orientation:***

Using this instrument, the student's gaze and head orientation are tracked, which provides greater insight into the student's focus and the issues that have grabbed their interest during the learning process.

- ***Spatial Precision:***

CNN specializes in very precise capturing of small local patterns in images. Such spatial precision is essential in measuring engagement levels, as it demonstrates the interaction or disengagement of the students from the specific areas.

In general, the CNN part, supported by the ResNet-50 architecture, has a major function in efficiently examining the engagement signals, which, through better comprehension of the student's behavior, leads to the improvement of educational experiences.

4.2.3. Vision Transformer (ViT) for Contextual Feature Learning

Vision Transformer VIT is a cutting-edge model architecture that is uniquely tailored to identify and comprehend complex, deep relationships with the facial features of a picture. Unlike traditional CNNs, which are mainly effective for structured grid-like data, a ViT applies a transformer architecture that makes it possible to handle the face images as sequences or sequences of patches.

The fundamental idea of the ViT is to extend the potential of CNNs by enabling the system to obtain both long-range dependencies and global spatial information that are usually missed by traditional CNNs. Consequently, the system gains a more comprehensive understanding of the less obvious spatial relationships that exist throughout the entire facial anatomy. The Vision Transformer functioning procedure takes several detailed steps:

- ***Patch Extraction:***

The very first step is to convert the feature map that the CNN has generated into patches of the same size. In other words, this process converts the 2D spatial representation of the face image into a sequence that is compatible with the transformer's input. Each patch is like a separate unit of the whole facial image; hence, the model can zoom in on the small-scale features.

- ***Tokenization:***

The patches are first extracted, and then they proceed to the next stage, which is called tokenization. By using a linear function, namely T, each patch is converted into a tokenized version. Tokenization essentially preserves the key spatial relationships in the patch and simultaneously reduces dimensionality, which is more convenient for the model to process.

- ***Self-Attention Mechanism:***

The tokenized patches are next fed through multi-head self-attention layers. This essential part of the model checks not only the similarities but also the differences of various facial features, even those that are at minor distances from each other. The self-attention mechanism enables the model to identify and prioritize those features that may be the least noticeable, but still bridge the nature of the subtle signals across the facial landscape.

- ***Feature Aggregation:***

The outputs from the attention layers are merged to obtain the one that reflects a more complete visual representation—a feature vector for the ViT. The feature vector thus not only merges the local information of the different patches but also provides the global context, which is like a bridge for a richer and more detailed understanding of the facial structure as a whole.

$$F_V = A(T(F_c, W_v))$$

Equation 12: Global Context Features Equation

Main Features of ViT:

- ***Long-Range Dependency Capture:*** The model is good at identifying and describing the intricate spatial relationships between various facial features that are necessary for activities such as face recognition and emotion detection.
- ***Context-Aware Learning:*** ViT can grasp the complete context of the facial movements by combining the information from the whole image, thus allowing a more profound understanding of the expressions and the gestures.
- ***Efficient Parallel Processing:*** The multi-head self-attention layers enable the model to handle data sequences efficiently from a computational point of view,

thereby making it suitable for working with vast datasets and intricate image representations.

To sum up, Vision Transformers represent a significant advancement in image processing, particularly for facial analysis, by leveraging intelligent design to achieve a more accurate and in-depth understanding of visual data.

4.2.4. Fusion Layer for Comprehensive Engagement Classification

The fusion layer is the primary connection that combines the local features, marked F_C, which were extracted from the CNN, with the global context features, F_V, obtained from the ViT component. Such a union results in features that are not only of high quality, but are also comprehensive enough to allow a subtle recognition of the levels of engagement, hence the most appropriate engagement classification from the different feature sets can be chosen.

- ***Feature Concatenation***

The feature embeddings from both the CNN and ViT modules are concatenated to start the process. By merging these two datasets into one vector, the model is able to depict the intricate and varied data that exists both in the local and global environments. Such a combined vector basically represents the data in the most complete way, thus enabling deeper analysis of the engagement behavior.

$$F_{fused} = [F_C, F_V]$$

Equation 13: Feature Concatenation Equation

- ***Fully Connected Layers***

After the concatenation, the combined vector is first passed through a series of fully connected layers, each of which uses a Rectified Linear Unit (ReLU) activation function. The layers serve to a great extent to the model in identifying the smallest and the largest changes in the context of the entities checked. By going through several rounds of the network's transformation, it essentially acquires and strengthens its complex patterns and relations with the given data, thus improving its capacity for making predictions.

- ***Attention-Based Weighting***

The attention mechanism, which controls the selection of features to focus on during training, is a main element of this design. This function permanently enables the model to adjust the extent to which different features are considered important; hence, one or more features may influence the result to a greater extent, while the model is still examining the entire set of data. Consequently, this attention-based method minimizes information loss, preserving vital features and disregarding irrelevant noise.

$$F_{attn} = A(W_a . F_{fused})$$

Equation 14: The Attention Mechanism Equation

where $\mathbf{W_a}$ are the learnable attention weights.

- ***Regularization Techniques***

To address the problem of overfitting, a common challenge in deep learning, two main regularization methods—batch normalization and dropout—are typically used. Batch normalization is a key component that helps a neural network train faster and more stably. It essentially standardizes the inputs to each layer, which leads to quicker convergence and increased stability. Conversely, dropout is a technique where a random set of neurons is temporarily disabled during each training cycle. As a result, the network must identify more universal and therefore more robust features. This improves the model's ability to generalize performance when tested on new data.

- ***Output Classification***

Ultimately, the output vector from the previous stages is converted by a Softmax function. This function rates the different engagement categories, i.e., "Engaged," "Neutral," and "Distracted", with probabilities. Such a probability-based output enables the system to provide clear-cut decisions as well as to indicate the degree of confidence for each engagement class, thus making the system more user-friendly and applicable in real-life scenarios.

$$P(y|X) = Softmax\ (W_o . F_{attn} + b)$$

Equation 14: The Output Classification Equation

where $\mathbf{W_o}$ are the final layer weights and **b** is the bias term.

Essentially, the proposed hybrid CNN–ViT model combines efficiently the local pattern recognition capability of CNN with the global context understanding power of ViT to recognize student engagement in OCR environments. The method elaborates on organized preprocessing, modular feature extraction, as well as fusion concepts, and thus, a solid classification pipeline.

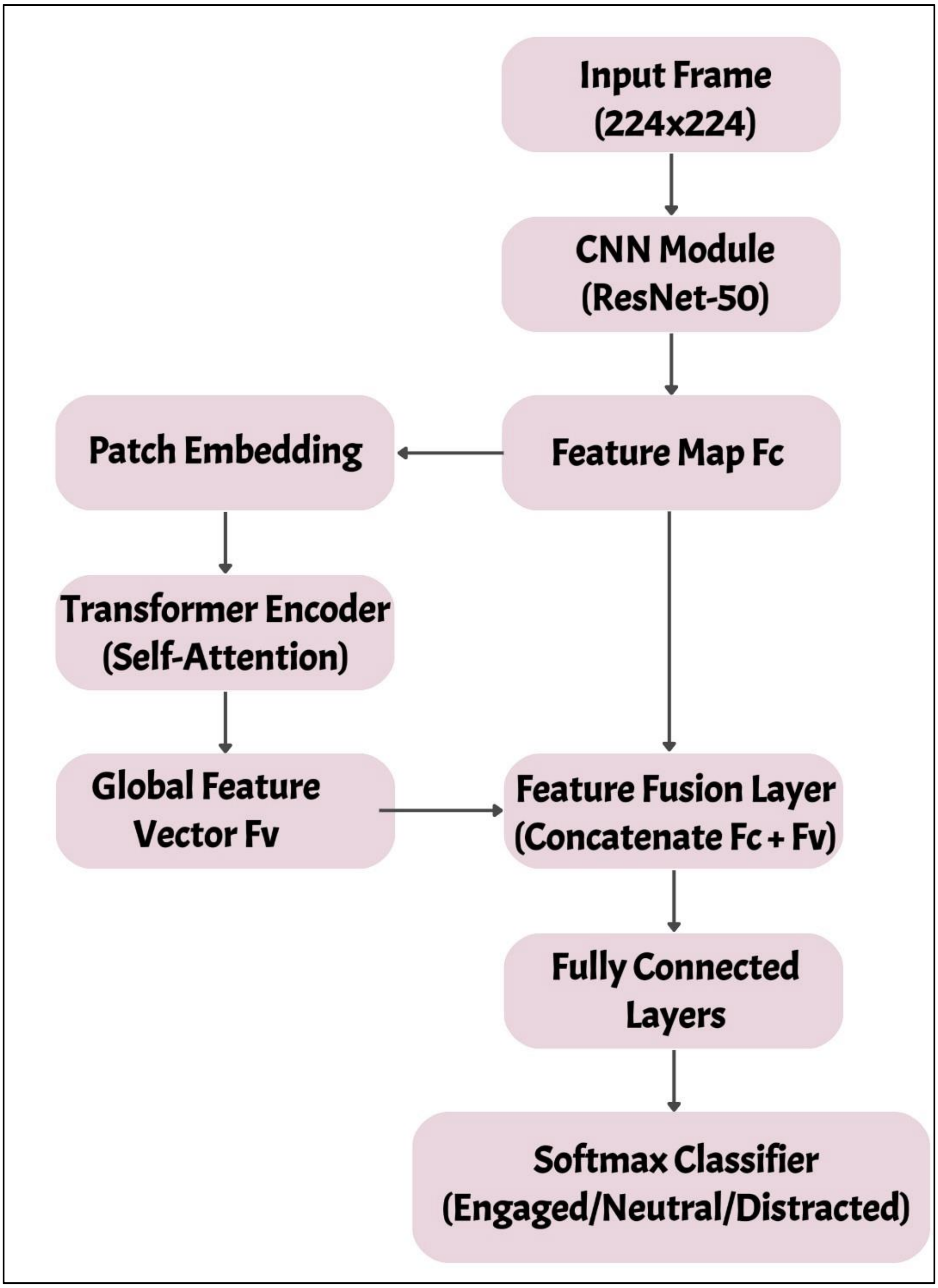

Figure 3-33: Hybrid CNN–ViT model Flowchart

4.2.5. Implementation Details

The development of the Hybrid CNN–ViT model to detect student engagement was essentially a meticulous design of hyperparameters aimed at achieving high generalization and performance across various learning settings.

- ***Training Configuration***

The batch size was set to 32, which was chosen to strike a balance between GPU memory consumption and the stability of gradient updates. The training limit for the model was set to 75 epochs. The early stopping procedure was led by the validation loss in order to control the overfitting of the model and keep its generalization ability.

The learning rate for the training was set to 0.0001, and a standard Adam optimizer, which utilizes adaptive momentum and exhibits fast convergence properties, was employed for the procedure. To provide additional safety for the network, a dropout of 0.5 was given to the fully connected layers, which, during training, randomly deactivated some neurons to allow the network to have more powerful features.

In the case of weight distribution, the method used was He normal distribution, which is especially effective for layers with ReLU activation functions as it speeds up the solution process and ensures stable gradients in the deeper parts of the CNN.

- ***Loss Function and Evaluation***

As this is a multi-class classification issue, the identification of the user's engagement states (Engaged, Neutral, Distracted), we made use of the categorical cross-entropy loss function. Such a loss function can be efficiently used for modeling the probabilities of different classes and providing a penalty for those incorrect confidence levels.

The performance of the designed system was measured by the values of accuracy, precision, recall, and F1-score that were calculated for the validation and test sets.

- ***Hyperparameter Selection Strategy***

The values of all hyperparameters were decided using a grid search technique, in which the different combinations of batch size, dropout, learning rate, and optimizer settings were tested. The best validation performance was used as a criterion to select the final configuration, which is in line with established best practices in deep learning for computer vision tasks.

Our CNN–ViT hybrid model has been properly trained to recognize the student engagement signal through the combination of well-tuned training settings, the use of a strong optimizer, and a judicious data partitioning strategy, which achieves an excellent performance-generalizability trade-off.

4.3. Experimental Results

In this section, we thoroughly analyze the effectiveness of the CNN-ViT model in detecting student engagement. The experiment and learning stages of our model are the first aspects we turn to for results. Next, we will compare the performance of our model with that of other existing models to display the particular abilities and benefits of the CNN-ViT hybrid approach. Our discoveries are a reflection of the model's precision, speed, and consistency.

4.3.1. Quantitative Evaluation of System Performance

4.3.1.1. Performance Outcomes: Analytical Discussion

This section demonstrates the proposed CNN-ViT performance in detecting students' engagement. We shall summarize the outcomes from the model execution phases, i.e., training and testing.

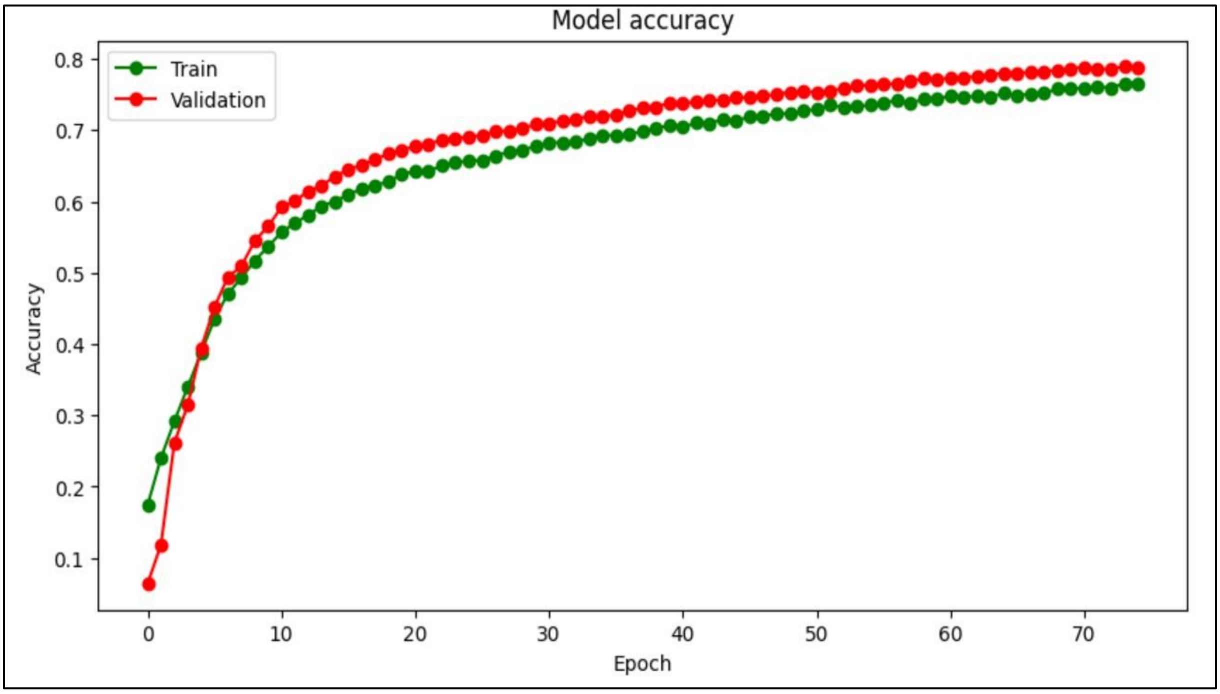

Figure 3-34: CNN-ViT Model's Accuracy

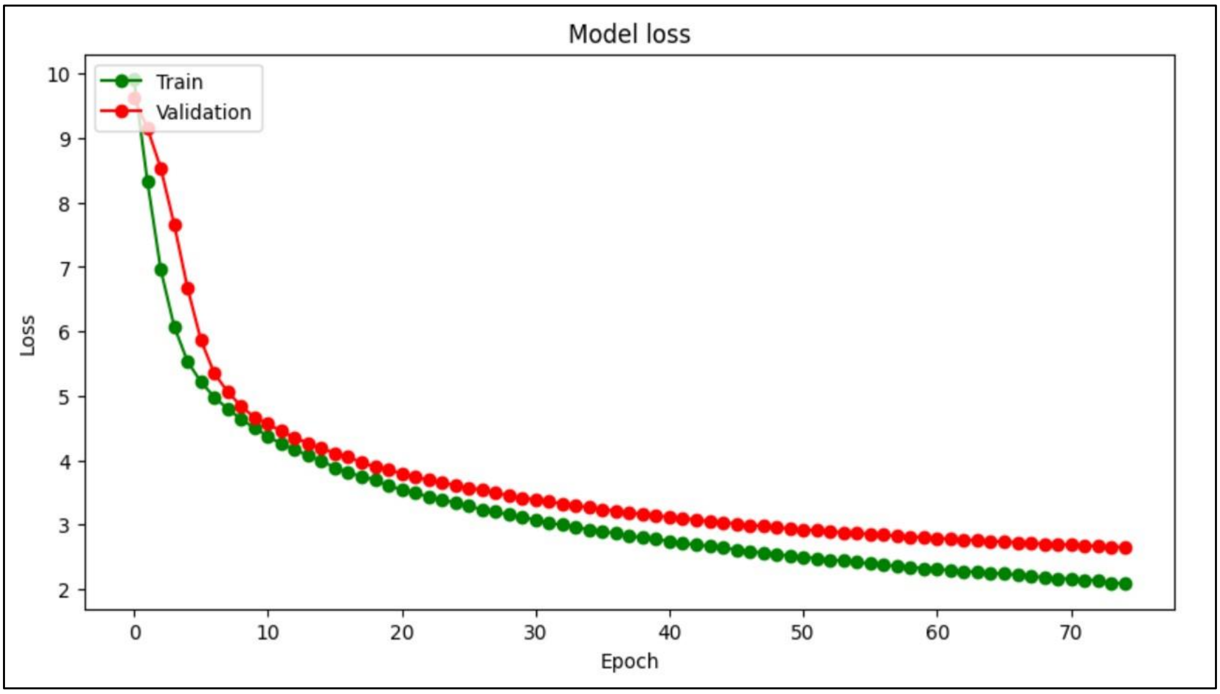

Figure 3-35: CNN-ViT Model's Loss

Essentially, the experimental results illustrated in Figure 3-34 demonstrate that the CNN-ViT model is stable and reliable when tested during both training and validation stages. To be more specific, the training accuracy trend can be seen to grow gradually over time. After the fifth epoch, the model achieved an average accuracy of 45%, and additionally, it reported a loss of 0.6. Yet, at the 75th epoch, the accuracy was almost four times higher and greatly varied from 75% to 80%. The model displayed a record of accuracy improvement, reaching 80%-85%.

The accuracy improvement strongly signifies that the model is indeed learning and gaining more skills to identify the features of the data related to student engagement. Over multiple training epochs, the model strengthened its knowledge of the particular characteristics that indicate student engagement, thus improving its prediction capability.

Additionally, throughout the training process, final loss values were consistently monitored, as shown in Figure 3-35, and they exhibited a decreasing trend, falling within the range of 0.2 to 0.4. This notable drop in loss indicates that the model has been effective in reducing classification errors during both training and testing. As a result, the alignment between the model's predictions and the true ground truth data has strengthened. This development suggests that the model has undergone successful training and is capable of accurately classifying inputs based on the features and patterns present in the dataset.

These findings support the original hypotheses about engagement detection activities, which identified the use of CNN and ViT architectures as the most effective. The CNN focuses on small facial features and body postures that are key indicators of engagement, while the ViT employs attention mechanisms for better understanding of these features. This combination not only forms the core of the student engagement classification system but also significantly enhances the model's performance during training and highlights the advantages of hybrid architectures in analyzing complex data.alysis.

4.3.1.2. Performance Metrics Evaluation

We relied on precision, recall, F1-score, and validation accuracy as our standard classification metrics to thoroughly measure the performance of the CNN–ViT hybrid model proposed. These metrics are most relevant in a multi-class classification setting, for instance, the student engagement detection problem, where the performance of all classes needs to be treated equally.

Table 3-10: CNN-ViT Model Evaluation Metrics

Metric	**Value (%)**
Precision	83
Recall	80
F1-Score	82
Validation Accuracy	80 – 85

Precision, at a high value of 83%, was the primary indicator of the model's ability to correctly select only the relevant instances of engagement (true positives among all predicted positives). Simultaneously, the recall at 80% indicates the model's completeness in identifying all actual instances for each engagement class, highlighting its receptiveness. The F1-score that links precision and recall through their harmonic mean was 82%, thus showing a good compromise between false positives and false negatives.

The model led to a validation accuracy ranging from 80% to 85%, indicating its good capacity for generalization to unknown data.

The diverse but mutually collective metrics provide a comprehensive examination of the model's performance. By inspecting the metrics for each class of engagement, we have unlocked the classifier's strengths and weaknesses. Such an exhaustive, metric-driven assessment not only attests to the CNN–ViT model's resilience but also underlines its applicability in practical educational surveillance situations.

4.3.2. Comparative Analysis with Existing Approaches

4.3.2.1. Cross-Method Performance Analysis

The comparative evaluation presented in Figure 3-36 allows in-depth analysis of various models for detecting student engagement. The comparison takes into account the training and validation accuracies as the main features of the model's learning effectiveness and its capability to generalize. Overall, the CNN–ViT hybrid model has the best performance and is superior to other baseline models.

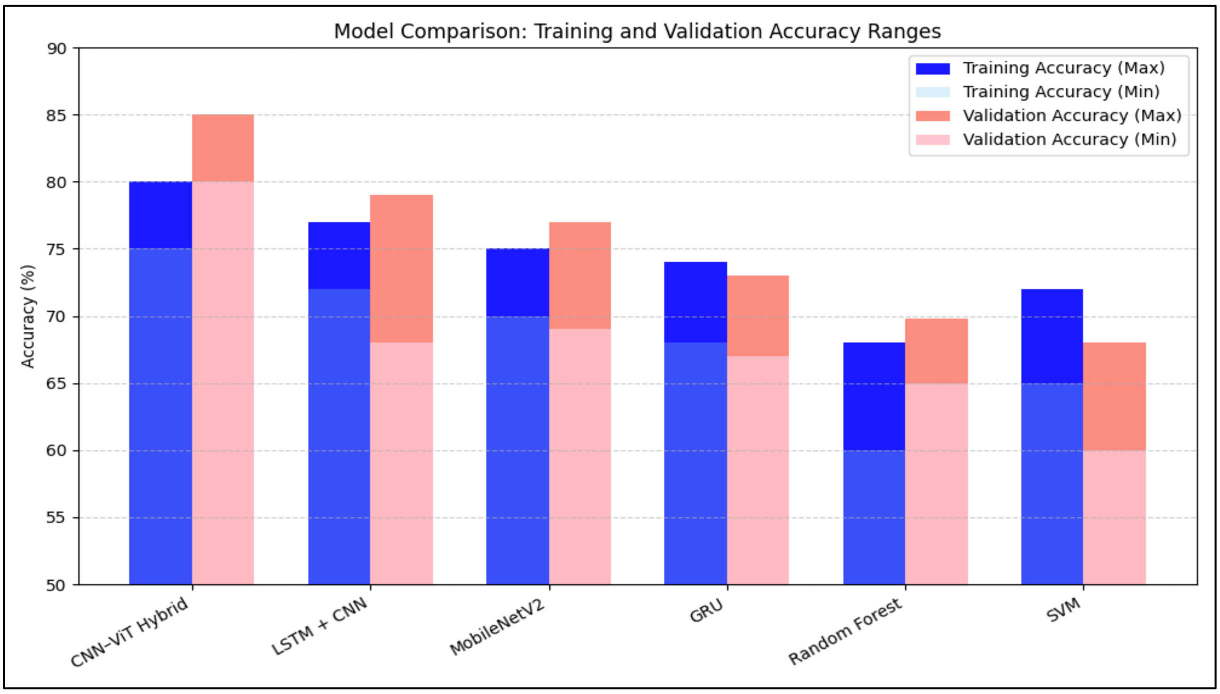

Figure 3-36: Model Comparison: Training and Validation Accuracy Ranges

The CNN–ViT hybrid model outperforms all other baselines by a large margin (Figure 3-36). It gets a training accuracy ranging from 75% to 80% and a validation accuracy estimated between 80% and 85%. This performance gap, where the validation accuracy is higher than the training, is a good sign of successful regularization and strong generalization. It signals the strength of utilizing the local spatial feature (CNN) together with the global contextual (ViT), which appears to be the winning combination to extract the behavioral changes stably.

The LSTM + CNN model is the second-best model considered. The training accuracy of this model is between 72% to 77% and the validation accuracy ranges from 68% to 79%. Its relatively high training performance suggests it can learn solid temporal patterns. However, the larger difference between training and validation results indicates that the model is sensitive to overfitting, especially in less-structured environments or when the dataset is noisy.

MobileNetV2 is a prime example of a model that optimizes the balance between performance and computational efficiency. It reaches 70%-75% of training accuracy and 69%-77% of validation accuracy. The design of this model allows it to be implemented in real-time on edge devices. However, it lags in deeper behavioral feature modeling.

GRU-based models show somewhat limited capabilities, where training accuracy is between 68% and 74% and validation ranges from 67% to 73%. GRUs are good at working with sequential data; however, they do not have the same level of detail as CNNs, which limits their performance when dealing with visual engagement.

Among the traditional machine learning models, random forests (RF) and support vector machines (SVM) are the worst performers. RF results in 60–68% training accuracy and 65–69.8% validation accuracy, while SVM achieves 65–72% training and 60–68% validation. The results indicate that classical models, despite their simplicity and interpretability, are not capable of representing the complexity of multimodal engagement signals in a dynamic classroom environment.

The CNN–ViT model generally outperforms the other methods during both the training and validation phases, highlighting its ability to extract and combine local and global visual features. Hence, the model can be considered adequate for detecting students' engagement, which is accurate, scalable, and can be performed in an OCR environment in real-time.

4.3.2.2. Evaluation of Selected Methods Based on Performance Metrics

The assessment presented in Table 3-11 provides a detailed examination of the six distinct models that have been utilized for identifying student engagement. Besides the various performance metrics, the evaluation also acknowledges the qualitative characteristics of these models, dealing with their implementation and usability.

Table 3-11: Comparison of Our Model with Other Approaches

Model	Training Accuracy	Validation Accuracy	Precision (%)	Recall (%)	F1-Score (%)	Scalability	Interpretability	Resource Consumption	Adaptability to Large Datasets
CNN-ViT Hybrid	75–80%	80–85%	83	80	82	Moderate	Moderate	High	High
LSTM + CNN	72–77%	68–72%	78	73	75	Low–Moderate	Low	High	Moderate
MobileNetV2	70–75%	65–70%	74	71	72	High	High	Low	Moderate
GRU	68–74%	63–70%	70	69	68	Low	Moderate	Moderate	Low
Random Forest	60–68%	55–65%	66	64	65	High	High	Very Low	Low
SVM	65–72%	60–68%	70	66	68	Low	High	Low	Low–Moderate

The assessment presents a detailed examination of the six distinct models that have been utilized for identifying student engagement. Besides the various performance metrics, the evaluation also acknowledges the qualitative characteristics of these models, dealing with their implementation and usability.

The CNN combined with LSTM is successful to a great extent, with the precision of 78% and F1-score of 75%; nevertheless, the accuracy of the validation set fluctuates between

68% and 72%. Such a signal can be interpreted as the model overfitting or being more vulnerable to the data's noise. On top of that, while this architecture can process the temporal features of student engagement, it still lags behind ViT-based models when it comes to harvesting the total global context from extended visual contexts.

The MobileNetV2 architecture achieved an F1-score of 72%, and it is known for its excellent efficiency and scalability; however, with a rather low total accuracy. A typical scenario for this model would be a resource-limited setting, such as an edge device or mobile learning platform, where stability and efficiency of operations are the primary priorities. Despite that, its low computational load is still maintained at the cost of both precision and recall being lower than those of the heavier models.

GRUs managed to perform at a somewhat average level, with an F1-score of 68% as a go-to metric. Typically, they are trained significantly faster and consume less computational power than LSTMs; however, they are less successful in tasks that require a thorough and profound understanding, such as deep visual behavior-based engagement analysis. As a result, this deficiency hinders their capability to dredge out the complex connection that could signify different levels of student engagement.

The prediction power of Random Forest and Support Vector Machine (SVM) models is lower than that of deep learning models, as their F1-scores are 65% and 68%, respectively. These conventional models fail to capture the complex characteristics of the high-dimensional, sequential, and visual features of the data, which are even necessary for correct engagement detection. However, they are very interpretable and have an easy deployment process, which makes them suitable for rapid prototyping or use in small-scale educational applications.

- ***Scalability:***

The use of lightweight models, such as MobileNetV2 and Random Forest, also their ability to be scaled, enables these models to be used on multiple hardware platforms without any problem. On the other hand, CNN–ViT, which is less scalable, still requires high computational resources due to its detailed attention mechanisms.

- ***Interpretability:***

Classic machine learning models (Random Forest and SVM) are notable for their interpretability, which is indispensable for upholding the transparent decision-making process.

Deep learning models negatively impact the transparency of the decision-making process to the extent that better prediction accuracies are achieved. Hence, they raise an issue for those users who need to understand the model's decision-making process.

- ***Resource Consumption:***

The LSTM + CNN and CNN–ViT models are extremely resource-intensive, particularly during the training process, and thus require a lot of computational power and memory. On the other hand, MobileNetV2 is a model specifically designed for cases when the computing resources are limited. Besides, Random Forest is the model that consumes the least amount of resources, respecting operational consumption.

- ***Adaptability to Large Datasets:***

The CNN-ViT model is the most versatile in the list due to its transformer backbone, which is capable of managing larger data with the same performance. The GRU and SVM models can not adjust to new situations as well as CNN–ViT because these have low adaptability, resulting from either the shallow architectures of their structural designs or non-parallelizable training algorithms, leading to limitations in their capability when large datasets are involved.

To sum up, the hybrid model CNN–ViT is the most effective method to recognize the involvement of students that combines symmetrically its power of prediction, handling large datasets, and practical use in the field. However, there may still be different kinds of models that are more suitable for specific cases, for example, those with very low computational costs or where interpretability, like MobileNetV2 or Random Forest, is highly valued. Whether to install which model first depends on the balancing of performance, specification of the issue, and deployment constraints that are compatible with the type of educational environment under consideration.

4.4. Discussion and Future Directions

The combination of CNN and ViT has demonstrated a strong capability to recognize student involvement in learning environments, achieving training accuracy of 75% to 80% and validation accuracy of 80% to 85%. The last loss values of the model are between 0.2 and 0.4, indicating effective learning and low error rates in predictions. Such performance levels are a good example of the compatibility of CNN and ViT.

CNNs excel in extracting detailed local features from images, which is necessary for detecting the slightest changes in a student's face, eyes, and posture. Whereas, ViTs are ideal for capturing long-range dependencies and contextual relations in the data by means of their special attention mechanisms. The combination of these two models offers a more thorough analytical framework that supports the recognition of subtle engagement signals, thus, even in difficult environmental settings with changing light, things blocking the view, or students with various behavior styles.

Such features lead to the stable performance over the different engagement categories, as shown by the even distribution of precision, recall, and F1-scores over different assessments. Consistency of this kind is essential as it shows that the model is not biased against any single engagement class. This attribute makes it particularly suitable for the dynamic OCR situation, where the students' attention can change a lot over time. Through the model, which automatically detects learners that may require support or intervention, personalized instruction and adaptive learning strategies can be developed, and the latter can be adjusted to the requirements of each student.

The CNN–ViT hybrid model is extremely limited in its practical application in the real world. ViTs need large amounts of computational power, which results in very high training costs and longer times of convergence, making it difficult to deploy on devices with low power and limited resources. Additionally, the complexity of the model also affects the interpretability, which makes the process of parameter fine-tuning and diagnostic analysis more difficult than that of simpler architectures like MobileNet or SVMs. Additionally, the achievement of this model still ties to having access to large, well-balanced datasets; thereby, most of the time, a lot of data augmentation or pretraining is required to be able to guarantee the effectiveness of the model.

Research plans to focus more on developing the CNN-ViT model optimization strategies in the future. Such measures might be model trimming, quantization, and knowledge distillation, all of which have a common goal of lowering the resource requirements for computation while keeping accuracy at a high level. These kinds of breakthroughs would be turning the model into one that can be operated in real-time on edge devices, such as tablets in classrooms. Moreover, we are looking into the various lightweight transformer models or the hybrid designs, which have fewer layers of attention to decrease the power usage even more without losing performance. Besides just the changes in architecture, utilizing transfer learning and domain adaptation methods will also be crucial for increasing the model's capability to recognize different cultures and educational contexts. Not only will this strategy guarantee the hold and spread of the model, but it will also make the model stronger when it comes to variations of classrooms worldwide.

To sum up, the CNN–ViT hybrid is a considerable step forward in the discovery of an automated system that can accurately detect student engagement and is not only high in accuracy but also robust against noisy inputs and capable of understanding both local and global behavioral cues. Applying upcoming optimization methods and performing cross-domain validation will allow this model to become a practical, on-the-spot solution for use in adaptive learning environments. The next work will aim at extending efficiency, scalability, and interpretability, thus becoming the basis for educational systems integration, allowing for deployment without losing strong predictive capabilities in environments with scarce resources.

5. *Contribution 4: HRT-GRU: A Hybrid Model for Student Progress Prediction in Open Classroom*

5.1. Description:

OCR environments represent a significant shift in modern education, emphasizing adaptability, student independence, and personalized instruction. This setting offers benefits such as improved access to materials, real-time feedback, and customizable learning paths. However, challenges remain; notably, the absence of standardized metrics for measuring engagement leads to inconsistent assessments. Additionally, the large volume of student activity data makes it hard to monitor individual progress, rendering traditional evaluation methods less effective and reducing the likelihood of timely, tailored interventions.

Predictive modeling is considered a viable solution to the issues mentioned. These models, by identifying students who are likely to perform below average early in the learning process, enable targeted support, adaptive learning strategies, and improved overall outcomes. However, most current methods mainly depend on static student characteristics, such as demographic information or isolated test scores, and fail to adequately capture the dynamic, hierarchical, and sequential nature of learning in OCR settings. This limitation highlights the need for more advanced AI models that can monitor students' behavioral development over time and account for the interconnected educational factors.

We propose an HRT-GRU prediction system to bridge the gap in predictive analytics within the education field. This system is an advanced hybrid deep learning model that effectively combines the Hierarchical Reasoning Transformer (HRT) modules with Gated Recurrent Unit (GRU) layers. Its goal is to explore not only the hierarchical relationships but also the time-dependent patterns in student data, providing educators with insights they can use to facilitate the learning process by implementing intervention strategies early.

This HRT system is powered by transformer attention mechanisms, which utilize hierarchical feature embeddings to be able to pull out not only structured but also unstructured educational information effectively. The HRT is how the program identifies the highest-impact aspects of the educational process by in-depth investigation of various factors, the results of assignments, the attendance rates, and the resource usage. This is a key feature that enables it to handle the complexities of factor interactions, revealing critical insights even in cases when these factors have complicated and non-linear relationships.

Conversely, the GRU part is mainly constructed to effectively represent the changes in time-dependent academic performance. This is done to allow the unit to keep only the relevant information from the past and get rid of the noise that makes it hard to see the progress. This double concentration lets the GRU be adjustably able to give a clearer view of students' performance paths.

Combining these two sophisticated modules not only allows the HRT-GRU system to benefit from the logical contextualization provided by transformers but also from the memory-based sequential learning of recurrent networks. The resulting model is a predictive system with the following important traits:

- ***High Accuracy:*** One of the major features of this model is high accuracy. It achieves this by incorporating both context and temporal changes, thus being more reliable and consistent.
- ***Computational Efficiency:*** The HRT-GRU is less computationally demanding than other recurrent models, such as the LSTM networks, because of its simpler architecture. This efficiency is reflected in the quicker processing times and lower resource requirements, thereby making it suitable for real-time applications.
- ***Flexibility and Interpretability:*** The framework is flexible, built to accommodate the different educational contexts by educators. Moreover, the system's understandable nature enables on-the-spot supervision and encourages practical feedback, which offers teachers easily accessible and usable in their instructional skills.

Such an advanced method enables educators to foresee possible learning problems that have not yet become serious; thus, they can not only increase student retention but also boost the general student performance, especially in OCR environments.

5.2. Methodology

This section illustrates the creative and practical aspects of a combined HRT-GRU model to forecast the educational growth of students in OCR settings. The motive behind creating such a design was to enable the handling of diverse data (for instance, participation, engagement, and academic records) and, at the same time, to extract the high-level data relationships and the time-dependent patterns of the learning process. The model functions on preprocessed data, which are logs of students' behavior and performance in both structured and semi-structured forms.

5.2.1. Dataset and Data Preprocessing

5.2.1.1. Dataset Description

In order to evaluate the efficiency of the HRT-GRU model, we utilized the Intelligent Classroom Dataset available on Kaggle [163]. Such a dataset, collected from smart classrooms equipped with multiple sensors and interactive devices, is a pioneering dataset for AI learning educational analysis. It is a big dataset with multiple features that show academic performance, behavioral signals, student engagement levels, and prevailing environmental conditions.

The dataset features thousands of entries that detail a student's learning sessions. Each session is enriched with various time-based indicators, including the attention span, emotional recognition scores, the number of interactions with peers and educators, test performance, and levels of participation in class activities. These measures provide a detailed portrayal of the educational engagement and other cognitive activities and performance outcomes in technology-supported learning environments, which represent the continuous changes that originate from learners' engagement.

- ***Engagement Metrics:***

These indicators cover such aspects as attention level, engagement index, participation score, and emotional state of the students, which depict the degree of students' active and emotional involvement in their learning activities.

- ***Academic Indicators:***

These are the scores of tests and quizzes conducted for assessing performance, historical data of student performance, and other academic-related metrics, providing a picture of students' academic trajectory and proficiency.

- ***Contextual and Environmental Data:***

Such features may comprise changes in sound level inside the class, the time of the day, and the position of the seat, accounting for the wider setting of learning and its possible effect on the level of engagement and performance.

- ***Demographic Attributes:***

These are the main demographic features of age, gender, and background that could provide an additional layer of understanding when studying the impact of demographic factors on the extent of learning.

The diversity and Comprehensiveness of the Intelligent Classroom Dataset are particularly suitable for training the HRT-GRU architecture's hierarchical attention modules and temporal modeling layers. The nature of the data facilitates a thorough exploration of the relationships among different variables in the context of students' learning experiences.

5.2.1.2. Preprocessing Pipeline

Due to the diverse and unstructured nature of the Intelligent Classroom Dataset, detailed preprocessing was necessary to facilitate effective training and ensure that the model's performance remains stable. The preprocessing workflow consisted of different crucial steps as mentioned below:

- ***Handling Missing Values and Outliers:***

- The process of locating empty and incomplete data entries was conducted with great care. Firstly, the entries considered irrecoverable have been deleted from the dataset.

- All non-numeric values were changed to NaN (Not a Number) by default, and after that, the missing cells were filled with median values of the corresponding features. The latter step was made to ensure that the dataset is not significantly affected by an extreme value.

- The extreme values in the dataset that could encompass anomalous occurrences of extremely high- or low-test scores were found by applying the interquartile range (IQR) filter. Next, these outliers were removed to prevent any potential bias in the model's learning process

- ***Feature Normalization:***

- To maintain the input scales that are consistent for all features and make the training process faster, "Min-Max scaling" was implemented.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Equation 15: Min-Max scaling

- With this normalization method, each feature is rescaled so that its values lie within [0, 1]. Such a change is essential to make the whole gradient-based learning efficient, since it ensures that the weight updates during training are proportional.

- ***Sequence Construction:***

- In order to depict the time-dependent features of the data, the dataset was changed to sequences of a specified window size, usually 20 consecutive sessions per student.

- Due to the overlapping time-series windows, the sequential nature of the learning process could be preserved, and more input data could be used for the GRU model, which consequently made the training more efficient.

- ***Categorical Encoding***

Categorical fields, such as emotional states, were handled by using suitable encoding methods. Variables in the dataset were encoded into numbers, which were either label encoding or one-hot encoding, depending on the variable and the specific requirement of the model.

- ***Train-Test Split:***

- The sequences were divided into training and testing sets after random splitting 80% of the data for training and 20% for testing.

- To keep the distribution of different classes balanced during the split, especially for categories like high vs. low performers, stratified splitting methods were employed. This step is crucial in ensuring the model's evaluation remains fair by providing a representative test set.

This all-inclusive data preprocessing step was essential for the model HRT-GRU to get data of top quality, data that was time-aligned and of hierarchical structure. The direct link between data quality and the model's effectiveness in learning and explainability, thus the results of the study, to a great extent.

5.2.2. Hierarchical Reasoning Transformer (HRT)

Conventional transformer models rely on a standard self-attention method that treats all input sequences identically. Although this method has been successful in many cases, it is not satisfactory when handling educational data, which is usually multi-dimensional and naturally hierarchically organized. The Hierarchical Reasoning Transformer (HRT) module broadens the scope of a regular transformer encoder by combining the hierarchical relationships as the interaction mechanism. This innovative layout involves three main stages, each of which is crucial for deepening the understanding of education data complexities.

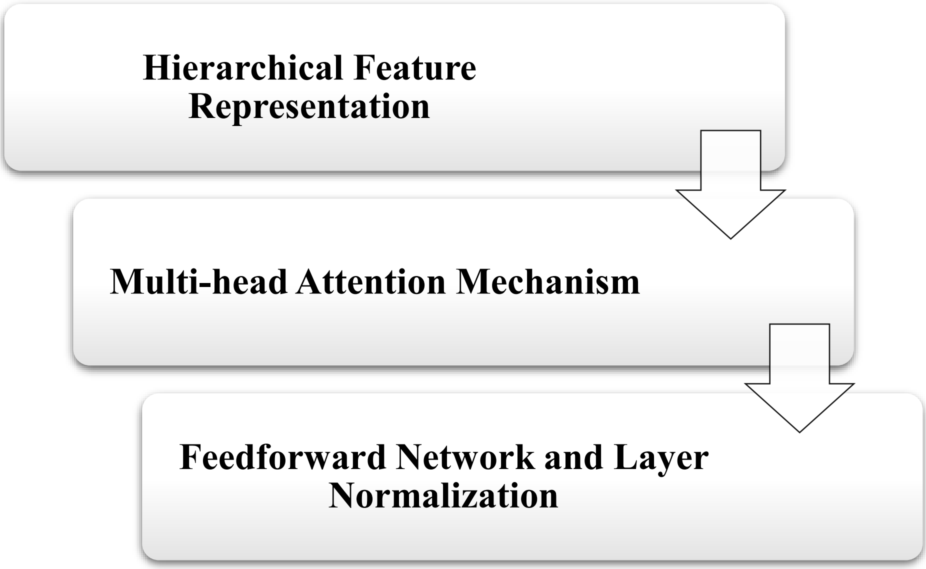

Figure 3-37: HRT module process

The HRT module follows a multi-step process illustrated in Figure 3-37:

- ***Hierarchical Feature Representation***

The educational dataset features various domains such as demographics, academics, and behavior. Demographic features include age and gender. Academic features can be test scores and class participation, while behavioral features may be attention levels and engagement indices. The HRT uses the hierarchical layer structure to represent these features and their relationships to each other.

For instance, the student's performance in mathematics based only on the latest quiz scores is not adequate. Recent scores can only be the outcome of the long academic trend of the student, consistent study habits, and active engagement with instructional materials over time. The HRT achieves this by allowing feature hierarchies to be nested, which in turn enables it to depict student learning pathways in such a detailed manner as to make them educational journeys.

- ***Multi-Head Attention Mechanism***

One of the crucial elements of the HRT unit is its multi-head attention system, which enables the model to focus on different parts of the input features simultaneously. Every attention head is separately trained to spot various relations in the provided information. Thus, as an example, one head might discover the association between history of grades and current performance, while another might identify the connections between students' emotions and their degree of engagement in the class. Besides these, a third head may consider how a student's practice of participation has affected their academic consistency.

$$Attention\ (Q, K, V) = \ softmax\ (\frac{QK^T}{\sqrt{d_k}})V$$

Equation 16: Multi-Head Attention Mechanism

Where:

- Q = Query matrix
- K = Key matrix
- V = Value matrix
- d_k = Dimensionality of key vectors

Such a thorough model of attention allows the model to find the features that are the most predictive of student success. The HRT can very effectively, by considering the same data patterns from different angles, give different weights to the various factors of a study, thus providing a more detailed account of the students' progress and requirements.

- ***Feedforward Network and Layer Normalization***

The multi-head attention mechanism output is further handled by the position-wise feedforward network and layer normalization stages. The feedforward network, which comprises multiple layers that employ non-linear transformations to represent implicit relationships in the data while keeping interpretability, is only partially used. Layer normalization plays a major role in the model's capacity to make the training process more stable and extend its generalization potential, as it controls the distribution of the activations in the network.

Dropout regularization is also inserted between layers to reduce the likelihood of overfitting. The principle behind this method is that a certain proportion of the input units is randomly chosen and set to zero during training, thus giving the model the necessity to learn stronger representations as it cannot rely only on a particular feature.

The HRT components, when viewed as one, enable the feature data to be converted to context-aware, hierarchically embedded representations. Such complex designs are necessary for the correct understanding of the student's behavior and capability; thus, the education system will be enhanced by the provision of precise insights and interventions.

5.2.3. Gated Recurrent Unit (GRU)

The GRU is basically a more sophisticated and improved version of RNNs, which typically have limited capabilities for complex dependency modeling in sequences. A prominent feature that sets GRU apart is the use of novel gates that cleverly regulate the flow of data through the network. This architecture, by design, addresses the issue of the disappearing gradient, which is at the root of most cases where standard RNNs perform badly, resulting in unstable learning processes and the difficulty of prolonging the training.

The design of the GRU enables it to retain loads of information for quite a long time, which is why it is suitable for time series tasks. One of the areas of the GRUs application is the educational data field, where they can monitor and forecast a learner’s educational development; additionally, they are also used in language processing, speech recognition, and predicting time-dependent future events from historical data.

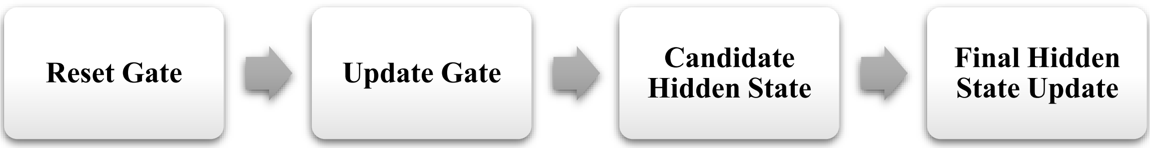

Figure 3-38: The Inner Mechanisms of a GRU.

A GRU inner mechanism is shown in Figure 3-38, which represents the detailed processes at every time step t. depicting the main functions outlined below:

- ***Reset Gate (rt):***

The reset power is important in selecting which part of the history the system will keep or discard. In simple terms, the reset gate gives the machine the capability of not being burdened with outdated or useless info by restricting the amount of past that can have an influence on the present. Hence, the network can still focus on the most relevant features of the input data, which is a very important condition for correctly handling long-term dependencies in time series.

$$r_t = \sigma(W_r x_t + U_z h_{t-1} + b_r)$$

Equation 17: Reset Gate Equation

- ***Update Gate (zt):***

The update gate, working as a dynamic controller, decides what fraction of the new input should be used for the hidden state. This gate, by bridging the retention of past knowledge and the introduction of new data, keeps the GRU not only relevant but also efficient with the latest data flows.

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

Equation 18: Update Gate Equation

- ***Candidate Hidden State ($\widetilde{\mathbf{h}_t}$):***

The candidate hidden state is generated from the current input and the last hidden state. serves as a new potential memory state that combines historical background with new enough data. This candidate is crucial as it offers a foundation for making informed updates to the hidden state.

$$\widetilde{h_t} = tanh(W_h x_t + U_h(r_t h_{t-1}) + b_h)$$

Equation 19: Candidate Hidden State Equation

- ***Final Hidden State Update ($\mathbf{h}_t$):***

The GRU changes its hidden state by mixing the new candidate hidden state and the previous hidden state in a weighted way, as the control is done by the update gate. This merging of information enables the model to combine memory with new input, which gives it the capability to be more accurate in its predictions and to make the right decisions.

$$h_t = (1 - z_t)\, h_{t-1} + z_t\, \widetilde{h_t}$$

Equation 20: Final Hidden State Update Equation

Where:

- x_t: Input at time *t* (e.g., student performance at a given assessment)
- h_t: Hidden state at time *t*
- W, U, b: Learnable parameters
- σ: Sigmoid activation function

In general, the GRU design is capable of effectively harnessing the features of sequential data, which thus makes it a versatile and potent instrument for various situations that refer to the prediction, treatment, and comprehension of time-dependent processes.

5.2.4. HRT-GRU Model:

The HRT-GRU model was developed with a modular design approach that allows each stage to be done separately and then integrated into the prediction pipeline. Such a procedure allows for each section to be critically tested and optimized, thereby ensuring the overall system's reliability and stability. The aim is to establish a robust predictive framework that can handle genuine classroom data, detect students' subtle learning behaviors, and provide educators with practical suggestions in an OCR environment.

The initial step to realize this goal was to convert the original data into a sequence-based format through extensive preprocessing. This change of format enables the data to be fed into the recurrent neural networks easily. Data cleaning, treating missing values, and encoding of categorical variables are the main activities conducted during the data preparation stage. Essentially, all these activities are to ensure that the models get high-quality inputs. After the data pre-processing stage, the model framework was developed using TensorFlow/Keras, which is a user-friendly environment for building deep learning models.

- ***HRT Block:*** Engineered with multi-head attention layers that were tailored for hierarchical input grouping.
- ***GRU Layer:*** A standard GRU unit implemented using Keras and is designed to handle sequence learning.

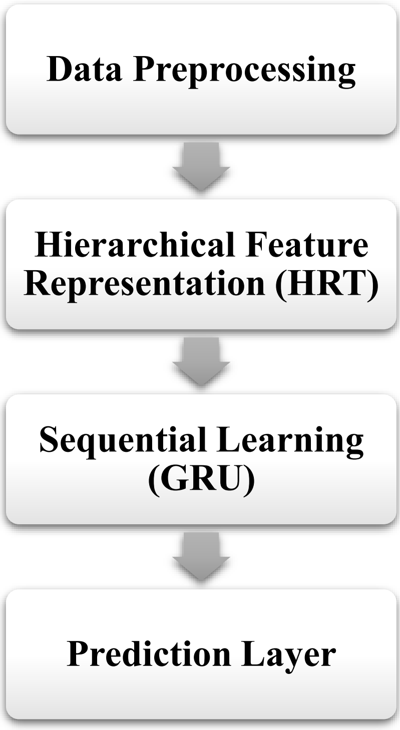

Figure 3-39: HRT-GRU model Workflow

Instead of being trained separately, the two units were combined one after another, the result of the HRT module being directly passed to the input of the GRU. In this way, the contextual and temporal patterns in the students' data were maintained.

This model implements the following steps in a careful and thought-out manner as illustrated in Figure 3-39.

- ***Input Shaping:***
  - The input sequences were shaped 3D tensors with the dimensions of (batch size, time steps, features), where time steps = 20. The sequences depicted the student's behavior data in a rolling window.
- ***Hierarchical Attention Encoding (HRT):***
  - Features were first aggregated based on their categories (for instance, academic, behavioral, and emotional).
  - Each category then went through the attention layers for encoding and was fed into dense sub-networks for further processing.
  - Next, the outputs were combined into a context-aware feature representation.

- ***Temporal Modeling (GRU):***
  - For handling the input data sequence, a GRU layer with 64 hidden units was employed to transform.
  - To mitigate the risk of overfitting, dropout regularization was applied.
  - The GRU output encapsulated the changes in the student context over the 20 time steps.
- ***Prediction Layer***

The prediction layer acts as a vital element of the HRT-GRU model, responsible for producing the final score indicating a student's performance. This stage employs a simple architecture with one neuron and a linear activation function, which is a perfect setting for regression problems. After the scores from the root interface and gateway layers have been collected, the system transforms them into a real number depicting the progress score of a student.

Mean Squared Error (MSE) has been used as the loss function for the model along with the AdamW optimizer. The performance of the model is maintained through the use of early stopping and batch normalization as stabilizers. The model, once made, can transfer the learnt weights to novel inputs, thus generating output that can be used for student tendency assessment. The effectiveness of the model is checked through different measures, for example, the R^2 value and accuracy rates, which basically aim to show a trace of student progress through their school-related data.

By deeply working with educational sequences, the system was designed to understand the context and predict performance changes over time using HRT and GRU. In addition, the "modularity" of the system means that it can also be adaptable for further developments, such as reinforcement learning loops.

5.2.5. Implementation Details

To improve the student progress forecast more effectively, the HRT-GRU was deliberately crafted to combine the context learning strength of the HRT with the time-dependent sequence processing abilities of the GRU.

The model works with input sequences of length 20, where each sequence consists of a set of normalized numerical features. As part of the preprocessing step, the dataset is subjected to MinMax normalization, which changes all the values of features to the range of 0–1. This step not only ensures that all attributes are on the same scale but also that the convergence will be faster.

- ***Transformer Component (HRT)***

The HRT module is designed using three consecutive encoder blocks. Each of these consists of:

- Multi-Head Attention: To extract the contextual dependencies among the different features.
- Feed-Forward Neural Networks: To change the representations in a nonlinear way.
- Layer Normalization: To ease the training process and also lessen the internal covariate shift.
- Dropout Layers: For regularization and avoiding overfitting.

Such a module is mainly the one to learn the dependencies, both hierarchical and cross-feature, exposed in student activity and performance data.

- ***GRU Component***

In order to represent the student's behavior changes over time, a single gated recurrent unit (GRU) layer has been added. This unit is able to hold onto previous states and develop skills to recognize the recurring patterns over time, which is essential to grasp the concepts of student engagement and performance growth.

- ***Dense Layers and Regularization***

After the outputs from the GRU and HRT, the model is equipped with fully connected dense layers:

- The nonlinearity is added through the use of the ReLU (Rectified Linear Unit) activation function.
- L2 regularization is included in the dense layers as a means of controlling the overfitting to the data. The method achieves this by applying a penalty to large weights of the network.

- ***Training Configuration***

- Optimizer: The model training process is accomplished with the AdamW optimizer, which essentially is a combination of adaptive learning and weight decay. The learning rate is set here at 0.001.
- Loss Function: The primary training loss is computed using Mean Squared Error (MSE).
- Evaluation Metrics: The model performance is evaluated quantitatively with Mean Absolute Error (MAE) and accuracy. These two metrics are also supported by the R^2 (coefficient of determination) score for the quality of regression.
- Early Stopping: An early stopping mechanism monitors the validation loss and interrupts the training if no improvement is found, hence, preventing overfitting and making the process time-efficient.

This setup enables the HRT-GRU model the ability to express features of the data with the right amount of regularization and training effectiveness, thus making it the perfect tool for a live educational progress monitoring system.

5.3. Experimental Results

The section details the experimental results of the innovative HRT-GRU model, which is designed to predict the behavior of students in OCR environments. After the model's training and testing results have been reported, we conduct an extensive comparative study, benchmarking the performance of the HRT-GRU model against several other models, looking at not only aspects such as model duration, training, and complexity, but also practical use. This kind of comparison uncovers the potential benefits as well as the drawbacks of the different strategies while also recognizing the contribution of the HRT-GRU model in the educational prediction analytics field.

5.3.1. Quantitative Evaluation of System Performance

5.3.1.1. Performance Outcomes: Analytical Discussion

The capacities of the HRT-GRU model proposed were showcased through its training and test plots for Mean Absolute Error (MAE) and loss over several epochs. These results provide a picture of the model's behavior in the training phase, its capability of generalizing to new data, and its stability.

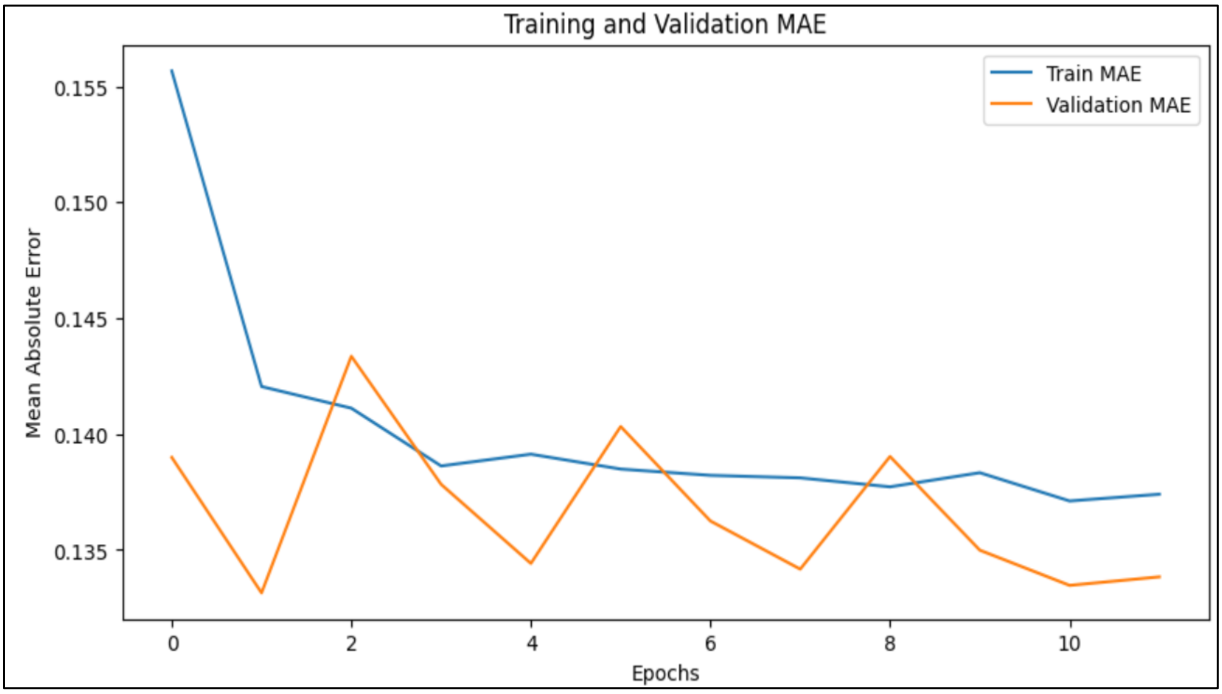

Figure 3-40: Training and Validation MAE results of 10 epochs

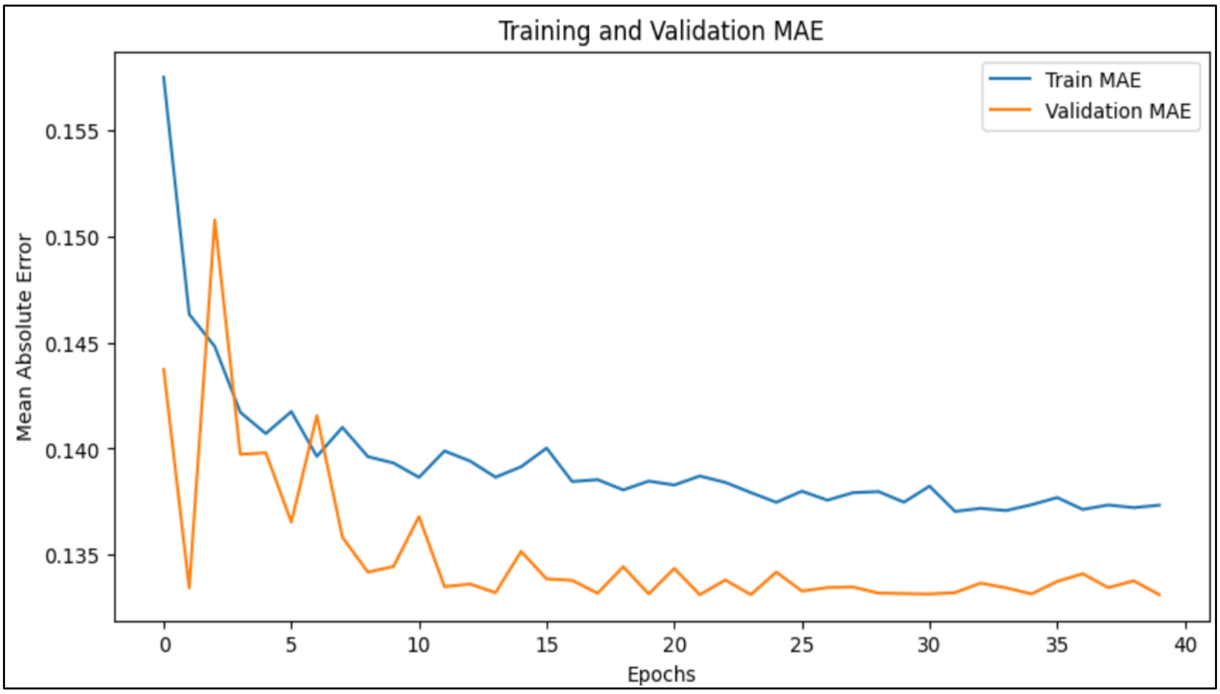

Figure 3-41: Training and Validation MAE results of 40 epochs

The first training process of the model, 10 epochs, as depicted in Figure 3-40, is marked with a notable reduction of the mean absolute error for both training and validation. The training MAE decreases significantly from around 0.156 to 0.135, indicating that the model has successfully adapted to the training data. At the same time, the validation MAE traces a similar decline, settling at values slightly lower than the training MAE towards the end of the epoch range. Such an instance implies that the model can extract hidden data patterns very quickly and still avoid overfitting, as confirmed by the small changes in the validation curve during the first iterations. The outcomes display that the model has low bias in its predictions at this point.

By extending the training to 40 epochs (depicted in Figure 3-41), the training and validation MAEs have almost doubled their progress, where the validation MAE acquired a minimum of 0.128. whereas the training MAE appears to have just under 0.140, thus, performance metrics have met their equilibrium. The closeness of these plots illustrates the large capability of the model to generalize, revealing its potential to make reliable predictions on new validation data.

The strong and clear decrease of both the training and validation MAE graphs, and also the very little difference between them, indicates that the HRT-GRU model is good at finding the right balance between bias and variance, thus giving a positive outcome. This equilibrium enables the model to make stable and accurate predictions over different datasets.

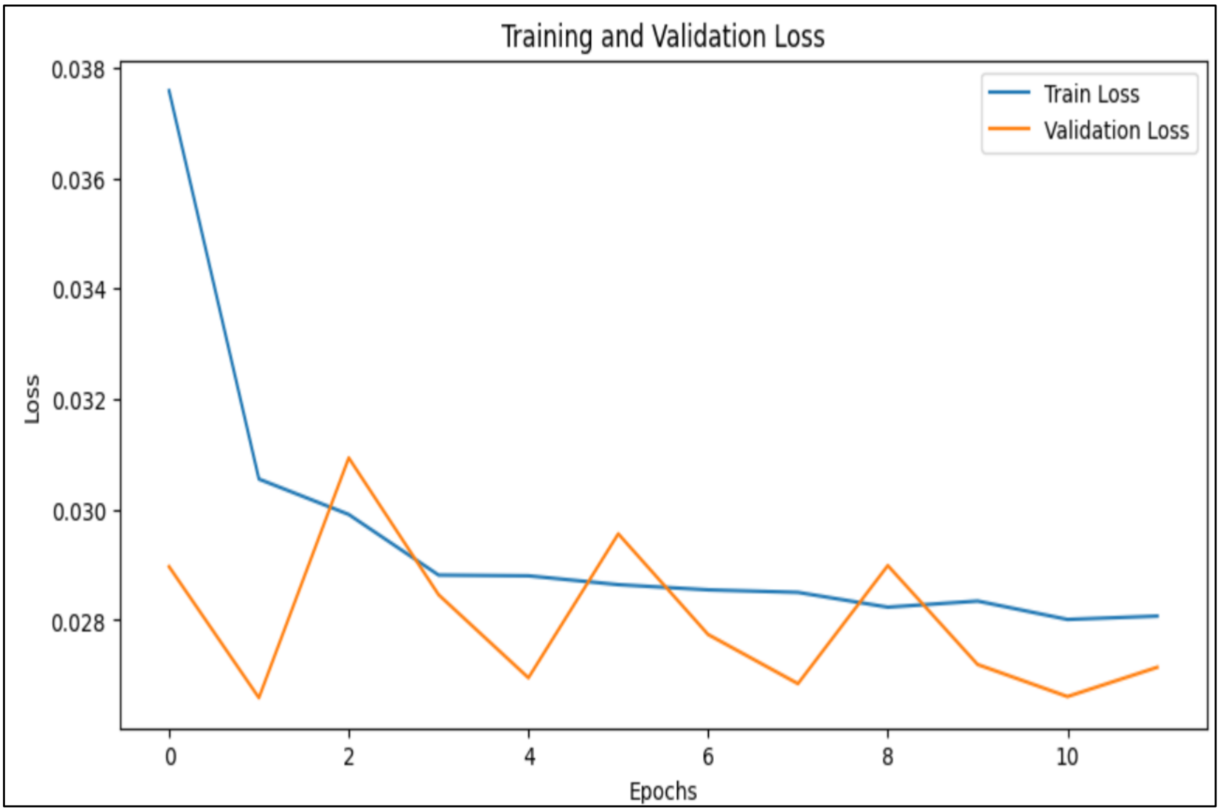

Figure 3-42: Training and Validation Loss of the 10 epochs

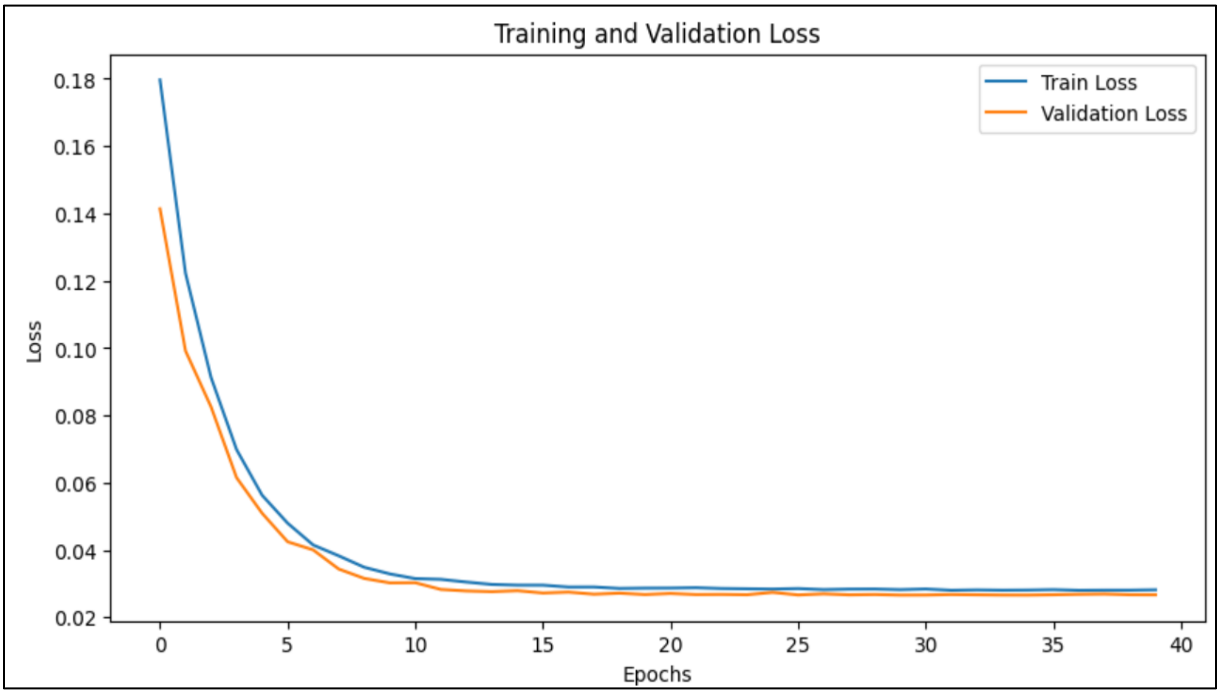

Figure 3-43: Training and Validation Loss of the 40 epochs

Figure 3-42 shows that the training loss for the first 10 epochs fell quickly in the first few iterations, dropping from about 0.038 to 0.028. After this initial drop, the loss continues to decline, but at a slower pace as training moves forward. Additionally, the validation loss follows this same pattern with minimal variations and is nearly the same as the training loss throughout the epochs. These hints suggest that the model is effectively identifying the relationships in the data, indicating an early convergence in the training process, with no signs of instability or erratic behavior.

Extending the training to 40 epochs (as shown in Figure 3-43), we notice that both training and validation losses decrease consistently. The loss is reduced from close to 0.18 to a little less than 0.04, and the validation loss is also dropping to similar levels. By the end of this training phase, both losses reach similarly low values, showing the model's strength. The similarity between the training and validation loss curves further supports that the model avoids overfitting. Improvements in the training dataset show in the validation dataset. The ongoing decline in loss metrics confirms that the model is strong and performs well, ensuring it works reliably during both training and validation.

The training and validation curves display a steady learning process, indicating effectiregularization. This stability comes from HRT's ability to find complex relationships in the learning data and GRU's skill in modeling time-based patterns and engagement trends. Together, these factors help the model detect subtle patterns in different educational datasets.

As a result, the hybrid structure reaches low error rates, high predictive accuracy, and reliable generalization.

5.3.1.2. *Performance Metrics Evaluation*

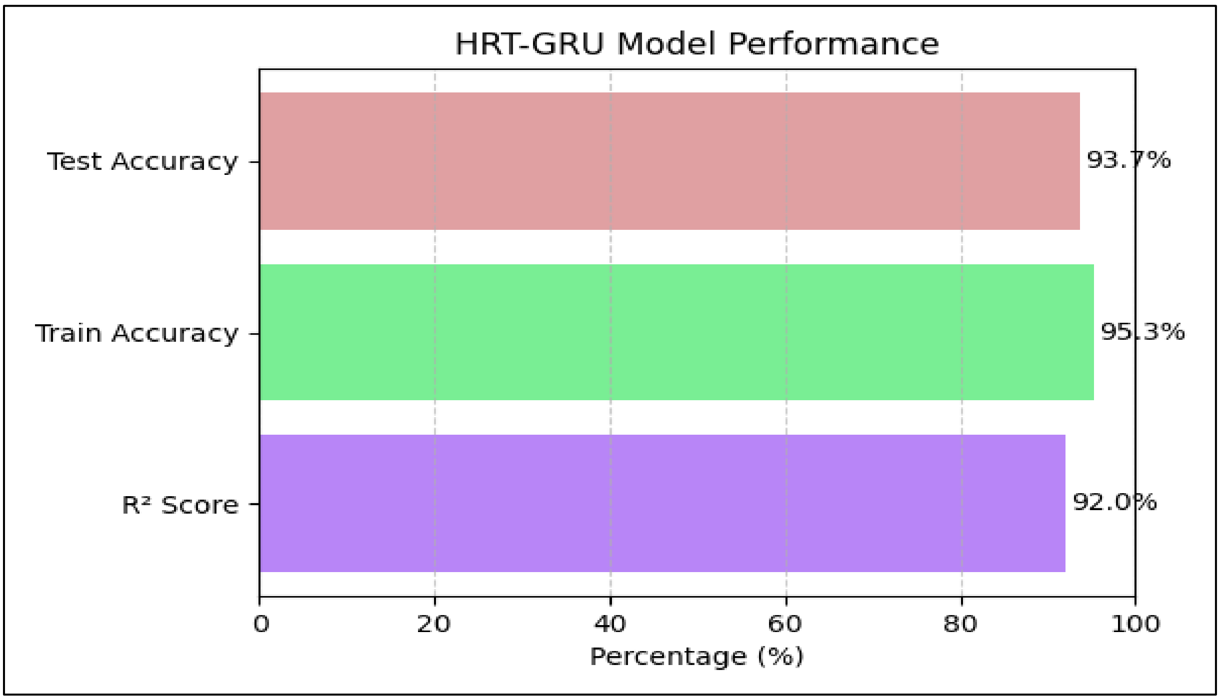

Figure 3-44: HRT-GRU Performance Metrics Evaluation

The performance evaluation data presented in Figure 3-44, strengthens the argument for the HRT-GRU model, which was presented as a hybrid model, standing out when compared with both classical and the latest baseline models. With this model, along with a training accuracy of 95.3%, the test accuracy of 93.7% is achieved, showing the generalization ability of the model to be quite close to the training one. This indirectly implies that it both recognizes the complex patterns of student progress data and refrains from overfitting, as it usually happens with machine learning models.

Moreover, the model has an elevated R^2 of 0.92, which is an indicator of its strong ability to make a good prediction. This means that the model is responsible for almost all the changes in the performance results of the students. High R^2 points to the fact that the model is successful in uncovering the real data features through close to accurate predictions.

The value of the model's MAE of 0.128 is crucial, as it is a very good example of the model's effectiveness in reality. The low MAE highlighted here means that the predicted performance scores are very often not far from the actual observed results, which is very crucial for teachers to be able to carry out accurate and timely interventions. Such an accuracy level not only deepens the educational indications from the model but also facilitates a pre-emptive approach in student performance management. In this case, the educators, by using the HRT-GRU model, can create personalized learning plans that truly match students' needs.

5.3.2. Comparative Analysis with Existing Approaches

5.3.2.1. Cross-Method Performance Analysis

Comparative assessment of prediction models presented in the Table 3-12 highlights that the HRT-GRU architecture leads all the other variants by a significant amount when the accuracy of the students' engagement in the open class context is considered.

Table 3-12: Model Performance Comparison

Model	R^2 Score	Train Accuracy (%)	Test Accuracy (%)	MAE
HRT-GRU	**0.92**	**95.3**	**93.7**	**0.128**
GRU	0.87	92.1	89.6	0.136
LSTM	0.89	93.4	91.2	0.155
Random Forest	0.78	88.3	85.9	0.189
SVM Regressor	0.71	84.6	81.4	0.205

The advantage of the HRT-GRU design over that of sequential traditional models like GRU and LSTM becomes more evident when such models are compared by their performance. Although LSTM achieved a relatively solid R^2 score of 0.89 and GRU a 0.87, both show higher MAE values (0.155 and 0.136, respectively), indicating less precision. The main reason for the superiority of our model over the rest is the presence of the HRT module, as it can capture the inter-relation of multi-layered context among data, while those that are purely recurrent fail such as GRU and LSTM.

Random Forest and SVM Regressor, which are the most common traditional machine learning methods, demonstrate significantly lower R^2 scores (0.78 and 0.71, respectively) and increased MAE values (0.189 and 0.205), thus indicating their inability to capture complex temporal and hierarchical relationships in educational data.

Generally, the outcomes support the concept of combining hierarchical reasoning with gated sequential learning as the core design. One of the main advantages of the HRT-GRU model is that it not only provides the highest precision and the lowest error but also stabilizes the performance of the system across the different stages of training and testing, thereby making it a very dependable tool for monitoring students' progress in real-time. Thus, teachers are allowed to recognize learners who are in danger of falling behind much earlier and create personalized, targeted interventions that are the best fit for them—that is, retention and success rates are improved as a result.

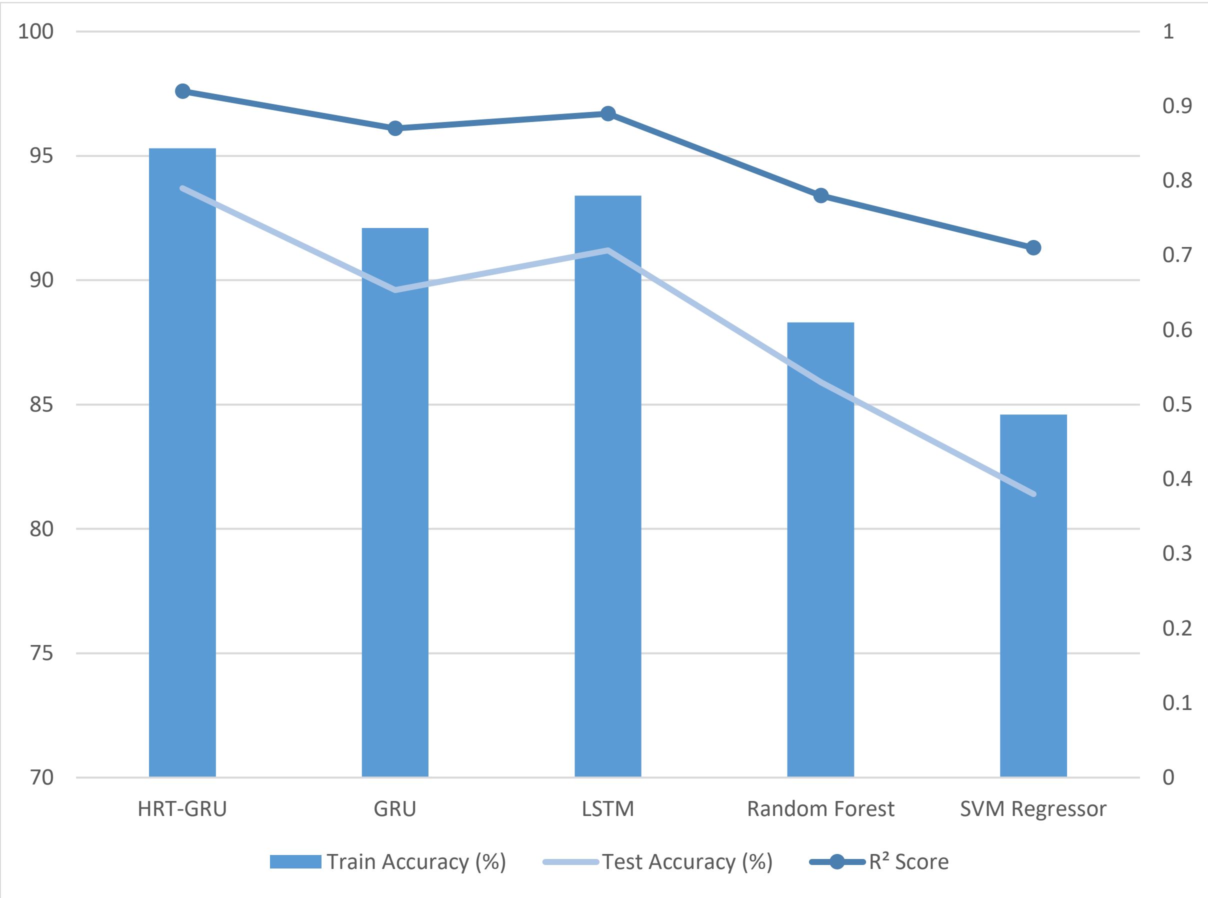

Figure 3-45: Model Performance Comparison Chart

5.3.2.2. Evaluation of Selected Methods

Table 3-13 summarizes the main features of the Intelligent Classroom Dataset and lists the expected effects of the features on student performance. These expectations were based on the results of behavioral modeling and by domain experts. The table presents the main features of the dataset, gives a short description of each, and explains the potential impact of each feature on the student learning outcomes.

Table 3-13: Influential Features on Student Performance

Feature	Description	Impact on Model Prediction	Practical Implication
Engagement Index	Monitoring students' focus and concentration during the session	High positive influence	Highly active students show a regular development of their skills
Attention Level	Real-time attention span measured with sensors	Moderate to high positive influence	A reduction of attention usually leads to a drop in performance
Previous Assessment	Current test scores or quiz results in school	Strong positive correlation with predicted performance	Helpful in identifying the patterns of learning
Participation Score	How often a student participates in class (e.g., hand-raising)	Moderate influence	Represents the desire of the learner to understand the material
Emotion Recognition	Predicted emotional condition from sensors (e.g., camera/audio)	Variable stressed or bored emotions lower performance	Contributes to the prevention of mental disorders at an early stage
Noise Level	Amount of background noise during study	Negative influence	Noisy classrooms negatively affect understanding and focus
Time of Day	Session time (morning, afternoon, evening)	Slight influence	Certain students learn best at particular times of the day

An analysis of the key features used in the HRT-GRU prediction system reveals that student progress forecasting is a complex problem that is influenced by various factors in the OCR setting. The Engagement Index is by far the most influential predictor among all the factors and is strongly positively correlated with the performance outcomes. Students who maintain high engagement levels are seen to improve academically in a measurable way, and this, in turn, reinforces the significance of facilitating interactive and immersive learning sessions.

Sensor-measured attention Level, captured in real-time, is among the important factors. The Positive Effect of a moderate-to-high nature indicates that performance declines are often preceded by attention dips, thus making it an excellent early-warning indicator. Moreover, the correlation between past and present academic performance is robust, endorsing that educational trends and previous learning achievements still hold a key to future academic success.

The Participation Score, which indicates the student engagement frequency during class, serves as a source of learner motivation and understanding. Despite its moderate influence, higher participation levels are often congruent with better learning outcomes, which means that active engagement strategies can be prioritized. The emotional state identified through Emotion Recognition has a significant but unique impact; negative emotions like boredom and anxiety are most times co-occurring with a decrease in the level of performance, thus stressing the importance of socio-emotional regulation and the delivery of support at the right time.

Environmental factors also influence the degree of learning. Noise Level negatively affects the abilities of a person, as a loud background noise can disrupt concentration and the process of understanding. Finally, Time of Day has a slight yet recognizable effect, thus suggesting that some learners could be more productive at certain times, a fact that could be utilized in the setting up of individualized schedules.

Collectively, these results indicate that the predictive efficiency of the HRT-GRU model is maximized when the model combines indicators of cognition, behavior, emotion, and environment. Because the factors are taken as a whole, teachers are in a position to plan specific treatments that penetrate through the educational inadequacies to the behavioral and environmental conditions that influence learning.

5.4. Discussion and Future Directions

The comparison of the HRT-GRU model to several baseline methods clearly shows that the model outperforms others in predicting students' performance in an OCR setting. With a squared correlation coefficient of 0.92, a training accuracy of 95.3%, and a testing accuracy of 93.7%, the model consistently exceeds the performance of the alternatives. GRU (R^2: 0.87, Test Accuracy: 89.6%), LSTM (R^2: 0.89, Test Accuracy: 91.2%), Random Forest (R^2: 0.78, Test Accuracy: 85.9%), and SVM Regressor (R^2: 0.71, Test Accuracy: 81.4%), all these models are outperformed by the HRT-GRU model. Using the MAE measure of 0.128 as an error metric indicates the model's potential to forecast student progress; it reflects its ability to produce reliable predictions while maintaining a very low margin of error.

Unlike single-architecture baselines, the hybrid design of HRT-GRU not only incorporates contextual reasoning through the Hierarchical Reasoning Transformer but also combines pattern temporal modeling via the GRU layer. This harmony allows one to concurrently identify the hierarchical relationships that exist within academic, behavioral, and emotional features, as well as the sequential learning patterns that alter over time. As a result, HRT-GRU could maintain balanced strengths in both domains, where GRU alone was strong in sequence learning but weak in capturing contextual nuance, and Transformer alone catered to context but lacked temporal sensitivity. Such a double feature was reflected in the training/validation curves, where both MAE and loss gradually decreased without overfitting, a performance that Random Forest and SVM, as weaker models, could hardly reach.

Upon further investigation of the Intelligent Classroom Dataset, it becomes clear that the predictive power of the model is due to different features. The Engagement Index and Previous Assessment scores were the biggest contributors; Attention Level and Participation Score also had significant roles in predicting academic performance. Emotion Recognition was a factor with both positive and negative effects, as negative emotions (stress, boredom) led to lower predicted scores. Noise Level was the main source of trouble. The effect of Time of Day was smaller but still interesting for possible personalized learning schedules.

On a practical level, the benefit that HRT-GRU holds over others is essentially its potential to uncover early warning signals like reduced engagement or increased noise times that are still far away from a significant performance decrease. An efficient use of this technology simply brings us the possibility of proactive interaction, emotional support, and personalized adaptation of learning, thus facilitating the educational progress of students.

The intentions for further research encompass broadening the current dataset to gather more varied behavioral and physiological data that could include voice tonality, body language, and eye movement; conducting more extensive testing of the model under different cultural and demographic classroom settings to ascertain its applicability in different scenarios; and, lastly, using the HRT-GRU model in actual teaching situations so that the educators can be notified at once if their assistance is required.

The experimental results show that HRT-GRU is an effective and balanced prediction model for smart classrooms. Going beyond typical and deep learning baselines, the model achieves higher precision and robustness, and simultaneously, provides understandable clues about the student performance factors. Thus, with lots of developments such as dataset enrichment, real-time deployment, and explainability, this model will be the core of future adaptive learning systems, which, in turn, will allow the educators to make the decisions based on the data that are personalized and at the right time for an ongoing learning process.

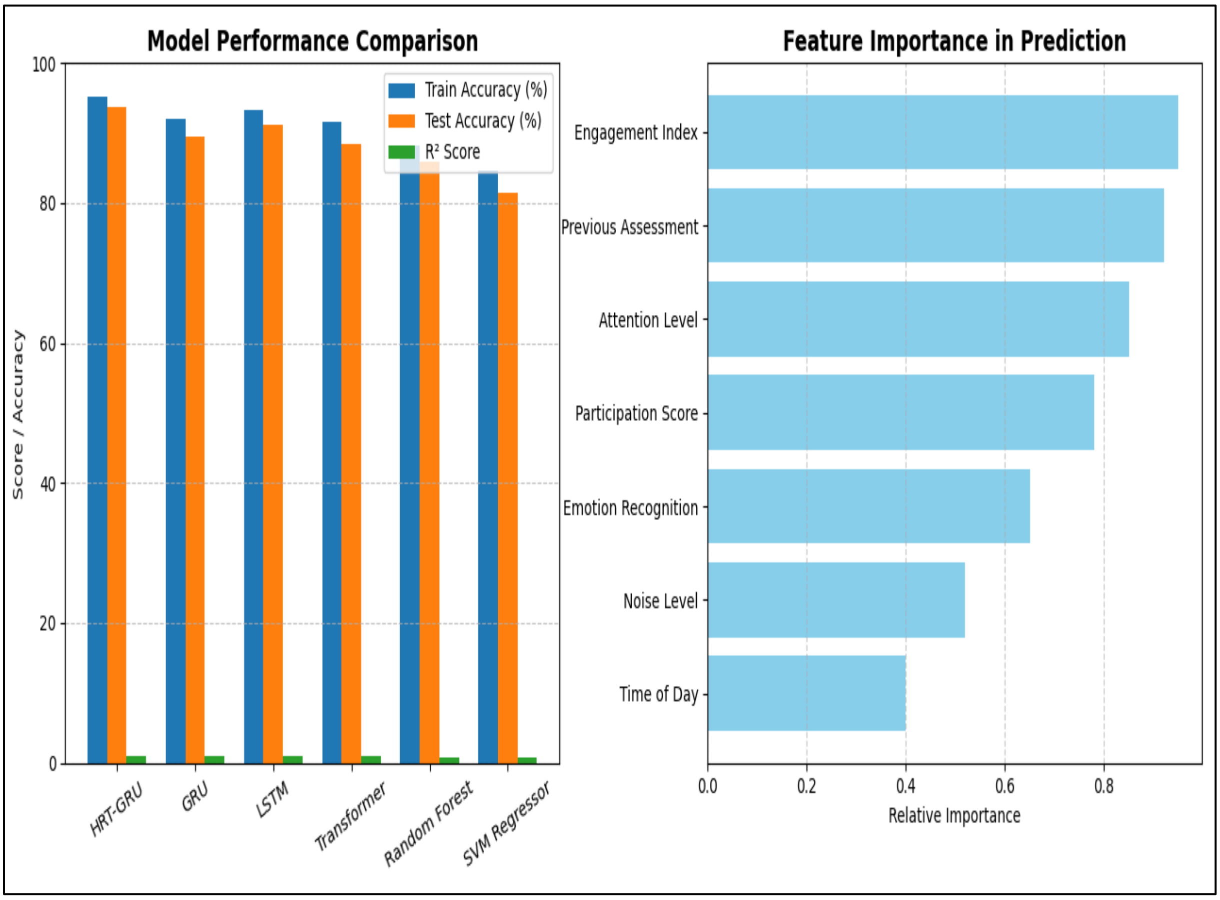

Figure 3-46: Summary (a Left) Comparison of Predictive Models in the Intelligent Classroom Environment (b Right) Key Feature Importance for Student Performance Prediction

6. Conclusion

With this chapter, we recap the four main contributions of this research. Each was presented as a separate solution to help solve the challenges related to the development of intelligent educational systems within the OCR environments. The LSTM-PSO-based intelligent search engine was the first one to be implemented. By combining deep sequential learning with particle swarm optimization, the developers aimed to increase not only the amount of data retrieval but also its relevance in any educational environment. Then the virtual assistant, EPO-T5-IoT, was developed, bringing along with it the perfect setting of a smart classroom where help is always available, and that too, timely, adaptable, and context-aware. The assistant is equipped with advanced NLP skills combined with EPO and real-time IoT integration for seamless coordination and better performance. Thirdly, we proposed a CNN-ViT hybrid framework for the detection of student engagement, this being the most accurate way, by combining the extraction of local features with the modeling of long-range dependencies. Lastly, the HRT-GRU model was set up to facilitate the prediction of student progress by combining hierarchical reasoning with temporal sequence modeling.

By combining deep learning, optimization, and IoT technologies, these contributions collectively show that it is both possible and effective to turn the Open Classroom into a learning space that is truly responsive, adaptive, and student-centered. The use of predictive models, such as progress monitoring, engagement detection, intelligent assistance, and enhanced search and retrieval, makes the learning environment capable of meeting students' changing needs in real time. As a result, instructors get the opportunity to intervene timely through insights, students receive personalized feedback and resources that enhance their learning, and the whole ecosystem is on the path to data-driven decision-making, which is retention, engagement, and overall academic success.

Experimentation and comparative analysis have been conducted to validate each solution, showing that each one performs better than existing methods by measurable metrics. The collaboration between AI models and Open Classroom paradigms is a practical and scalable model for the educational system of the future that is technologically enhanced but still relies on human interaction for teaching.

In the next chapter, General Conclusion, we will recapitulate the main points of the present study, discuss the research objectives, and review each contribution. Additionally, we shall address the limitations concerning data collection, methodology, and extent. Finally, propositions further studies, offering novel questions, combining different disciplines, and utilizing the latest technological advancements for achieving better learning outcomes in AI-based open classroom systems.

GENERAL CONCLUSION

This research aimed to tackle various problems related to eLearning, as well as to improve the Open Classroom model by using technological advancements in the field of artificial intelligence. We outlined the gap in academic research by pointing out the deficiencies of the current systems, which in turn enabled us to establish eLearning as the core of modern education, and thus, justify our work. The Open Classroom was introduced as an idea with great potential but still needing further development, especially with the help of smart and data-driven tools to facilitate its implementation.

The dissertation was organized clearly and coherently. It first placed eLearning and the Open Classroom in context, then it compared the work with other relevant research, and later it moved to the creation and assessment of the suggested contributions. Such an order allowed the features to be not only strong from a technical point of view but also educationally and user-wise meaningful. The findings revealed significant advances, confirming the use of AI as an innovative tool that could greatly facilitate the implementation of the Open Classroom concept and further foster the adoption of the Open Classroom vision.

To overcome these issues, we developed multiple creative AI-powered solutions that complement each other:

The LSTM-PSO intelligent search engine was the first system proposed to contextualize user queries for more relevant and accurate results, thus altering the entire retrieval process through radical innovation. The use of particle swarm optimization (PSO) combined with Long Short-Term Memory (LSTM) networks achieved a retrieval performance and relevance of approximately 95%, which is more than that of conventional LSTM-based systems. This improvement allowed simplifying the educational resource acquisition process and ensuring that those resources were the closest ones to the users' learning contexts.

The T5-EPO-IoT virtual assistant, the second approach, was another step forward, combining transformer-based natural language processing (NLP) with metaheuristic optimization algorithms, as well as with complete IoT functionalities. Results were a 95% improvement in accuracy, flexibility, and task execution over baseline T5 implementations. The system aims to create a smart tutoring system by providing effective answers to students' questions and creating more suitable learning paths.

The CNN-ViT model merged the distinctive features of both convolutional neural networks (CNNs) and vision transformers (ViTs) to identify the students' engagement. The model underwent the training phase with an accuracy between 75 and 80%, and was validated with an accuracy of 80-85%. In addition, the final loss was continuously between 0.2 and 0.4. This model allowed the detection of students' interaction levels, thus ensuring the early and accurate allocation of assistance.

Finally, the HRT-GRU model effectively integrated the basic concepts of hierarchical processing with the features of gated recurrent units (GRUs) to predict the involvement of students. This particular model achieved a high accuracy level ranging from 90 to 93%, which meant a considerable step forward in the area of adaptive assessment. The development of these types of predictions enables educators to customize the interventions that will have the greatest positive impact on students' progress.

These four systems, when combined, deeply enhanced the Open Classroom structure, which in essence is the main idea of customization, activity, and learner-centered design. Consequently, the Open Classroom model turned a practical setting with instantaneous adaptation to students' needs and preferences instead of just being a theoretical framework. Besides, this study highlighted the prospect of combining AI-based decision-making and IoT functionalities, transforming the education process from simple content consumption to a more engaging and participatory one.

Overall, the outcomes of this dissertation significantly improve the e-learning landscape. Traditional educational systems have consistently faced issues, such as static content delivery, insufficient personalization, and limited ways to monitor student engagement. This study demonstrated that combining deep learning architectures with optimization algorithms and IoT functionalities can effectively address these problems. The enhanced Open Classroom Initiative has the potential to become the next generation of flexible, responsive eLearning systems that adapt to various learning scenarios, ultimately bridging the gap between educational goals and data-driven insights.

The future scope of the research: several paths for future research remain open and interesting. The models employed achieved positive results in their predictions; however, there are issues of scalability, interoperability with currently existing educational systems, and significant ethical aspects, such as data privacy [164], [165], security[166], [167] and fairness [165], that have to be solved systematically before any wider application can be envisaged.

Moreover, the next phase of research may incorporate the implementation of reinforcement learning methods for continual learning, delving into multimodal data streams that integrate text, sound, image, and physiological signals for attention recognition, as well as further exploiting IoT connectivity to create highly engaging smart classrooms. Furthermore, a comprehensive, protracted study carried out in different educational institutions would provide essential information about the extent to which such solutions could be generalized and remain stable over time.

As a summary, the work of this dissertation made a remarkable contribution to closing the gap between the restrictions of traditional eLearning environments and the ambitious concept of a completely adaptive Open Classroom. Through the creation of the AI- and IoT-powered smart systems, it not only solved the inherent problems of the areas mentioned above but also opened up a more futuristic insight for intelligent, data-driven, and human-centered educational environments. The research put forth signifies a progressive move towards redesigning learning as a vibrant, captivating, and individualized experience that can adjust to the user's latest needs in the digital world, thus being adaptive.

References

[1] B. S. Oumaima, B. Abdemadjid, M. Majda, R. Mounira, and M. Derdour, ‘Examining Intelligent Tutoring Systems and Their AI Underpinnings in the Design of the Future of Learning’, in *2025 International Conference on Networking and Advanced Systems (ICNAS)*, Oct. 2025, pp. 1–8. doi: 10.1109/ICNAS68168.2025.11298109.

[2] I. Boutabia, A. Benmachiche, A. A. Betouil, M. Boutassetta, and M. Derdour, ‘A Survey on AI Applications in the Open Classroom Approach’, in *2025 International Conference on Networking and Advanced Systems (ICNAS)*, Oct. 2025, pp. 1–7. doi: 10.1109/ICNAS68168.2025.11298090.

[3] ‘Elliott Masie’, masie.com. Accessed: Jan. 04, 2024. [Online]. Available: https://www.masie.com/elliott-masie

[4] ‘TechLearn 2025 Conference | Home’. Accessed: Aug. 18, 2025. [Online]. Available: https://www.techlearnconference.com/

[5] B. K. Sedraoui, A. Benmachiche, A. Makhlouf, and C. Chemam, ‘Intrusion Detection with deep learning: A literature review’, in *2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, IEEE, 2024, pp. 1–8. doi: https://doi.org/10.1109/PAIS62114.2024.10541191.

[6] D. Hawkridge and J. Robinson, *Organizing Educational Broadcasting*. Taylor & Francis, 2023.

[7] A. Watters, *Teaching Machines: The History of Personalized Learning*. MIT Press, 2023.

[8] B. Cope and M. Kalantzis, ‘A little history of e-learning: finding new ways to learn in the PLATO computer education system, 1959–1976’, *Hist. Educ.*, vol. 52, no. 6, pp. 905–936, Nov. 2023, doi: 10.1080/0046760X.2022.2141353.

[9] S. Zhussupbayev, S. Nurgaliyeva, N. Shayakhmet, G. Otepova, A. Karimova, B. Matayev, H. Bak, ‘The Effect of Using Computer Assisted Instruction Method in History Lessons on Students’ Success and Attitudes’, *Int. J. Educ. Math. Sci. Technol.*, vol. 11, pp. 424–439, Jan. 2023, doi: 10.46328/ijemst.3136.

[10] Y. Mingming, Z. Bolun, L. Zhijian, W. Yingli, and Z. Lanshu, ‘Effectiveness of computer-based training on post-stroke cognitive rehabilitation: A systematic review and meta-analysis’, *Neuropsychol. Rehabil.*, vol. 32, no. 3, pp. 481–497, Mar. 2022, doi: 10.1080/09602011.2020.1831555.

[11] M. A. Camilleri and A. C. Camilleri, ‘The Acceptance of Learning Management Systems and Video Conferencing Technologies: Lessons Learned from COVID-19’, *Technol. Knowl. Learn.*, vol. 27, no. 4, pp. 1311–1333, Dec. 2022, doi: 10.1007/s10758-021-09561-y.

[12] M. Vasan, ‘Impact of promotional marketing using Web 2.0 tools on purchase decision of Gen Z’, *Mater. Today Proc.*, vol. 81, pp. 273–276, Jan. 2023, doi: 10.1016/j.matpr.2021.03.188.

[13] S. O. Boufaida, A. Benmachiche, M. Derdour, M. Maatallah, M. S. Kahil, and M. C. Ghanem, ‘TSA-GRU: A Novel Hybrid Deep Learning Module for learner behavior analytics in MOOCs’, Jul. 24, 2025, *Preprints*: 2025072023. doi: 10.20944/preprints202507.2023.v1.

[14] L. Q. Nguyen, ‘Learners’ satisfaction of courses on Coursera as a massive open online course platform: A case study’, *Front. Educ.*, vol. 7, Dec. 2022, doi: 10.3389/feduc.2022.1086170.

[15] P. C. Guyett, D. Chew, V. Azevedo, L. C. Blennerhassett, C. Rosca, and E. Tomlinson, ‘Optimizing SEM-EDX for fast, high-quality and non-destructive elemental analysis of glass’, *J. Anal. At. Spectrom.*, vol. 39, no. 10, pp. 2565–2579, 2024, doi: 10.1039/D4JA00212A.

[16] S. Dianati, ‘Behind the Screens: Unpacking the Ideologies of Udacity Through Ideographs and Criticism’, in *The Commercialisation of Massive Open Online Courses: Reading Ideologies in Between the Lines*, S. Dianati, Ed., Cham: Springer Nature Switzerland, 2024, pp. 87–105. doi: 10.1007/978-3-031-58184-7_6.

[17] A. Benmachiche, A. Sahia, S. O. Boufaida, K. Rais, M. Derdour, and F. Maazouzi, ‘Enhancing learning recommendations in MOOC search engines through named entity recognition’, *Educ. Inf. Technol.*, 2025, doi: 10.1007/s10639-024-13308-4.

[18] S. O. Boufaida, A. Benmachiche, M. Maatallah, and C. Chemam, ‘Hybrid Secure Routing in Mobile Ad-hoc Networks (MANETSs)’, Jan. 09, 2026, *arXiv*: arXiv:2602.13204. doi: 10.48550/arXiv.2602.13204.

[19] M. Boutassetta, A. Makhlouf, N. Messaoudi, A. Benmachiche, I. Boutabia, and M. Derdour, ‘Cyberattack Detection in Smart Cities Using AI: A literature review’, in *2025 International Conference on Networking and Advanced Systems (ICNAS)*, Oct. 2025, pp. 1–9. doi: 10.1109/ICNAS68168.2025.11298103.

[20] I. Boutabia, A. Benmachiche, A. Makhlouf, and M. Maatallah, *Recommendation system based on collaborative filtering in the library of El Tarf University*. 2021.

[21] J. Bojiah, ‘Effectiveness of Moodle in Teaching and Learning’, *J. Hunan Univ. Nat. Sci.*, vol. 49, no. 12, 2022, Accessed: Aug. 18, 2025. [Online]. Available: https://www.jonuns.com/index.php/journal/article/view/1259

[22] W. Aljedaani, M. Alkahtani, S. Ludi, M. W. Mkaouer, M. M. Eler, M. Kessentini, and A. Ouni, ‘The State of Accessibility in Blackboard: Survey and User Reviews Case Study’, in *Proceedings of the 20th International Web for All Conference*, in W4A ’23.

New York, NY, USA: Association for Computing Machinery, avril 2023, pp. 84–95. doi: 10.1145/3587281.3587291.

[23] P. Chatterjee, R. Bose, S. Banerjee, and S. Roy, ‘Enhancing Data Security of Cloud Based LMS’, *Wirel. Pers. Commun.*, vol. 130, no. 2, pp. 1123–1139, May 2023, doi: 10.1007/s11277-023-10323-5.

[24] P. L. S. Barbosa, R. A. F. do Carmo, J. P. P. Gomes, and W. Viana, ‘Adaptive learning in computer science education: A scoping review’, *Educ. Inf. Technol.*, vol. 29, no. 8, pp. 9139–9188, Jun. 2024, doi: 10.1007/s10639-023-12066-z.

[25] C.-Y. Hsu, I. Horikoshi, H. Li, R. Majumdar, and H. Ogata, ‘Extracting Students’ Self-Regulation Strategies in an Online Extensive Reading Environment using the Experience API (xAPI)’, *Int. Conf. Comput. Educ.*, Nov. 2022, Accessed: Aug. 20, 2025. [Online]. Available: https://library.apsce.net/index.php/ICCE/article/view/4501

[26] M. Majda, S. Hassina, B. Abdelmadjid, and M. Amina, ‘Un système de recommandation hybride basé flou adapté à l’apprentissage amélioré par la technologie’, *المجلة الجزائرية للعلوم*, vol. 1, no. 2, pp. 57–64, Jun. 2018.

[27] M. Boutassetta, A. Makhlouf, N. Messaoudi, A. Benmachiche, and I. Boutabia, ‘Hybrid IDS Using Signature-Based and Anomaly-Based Detection’, Jan. 17, 2026, *arXiv*: arXiv:2601.11998. doi: 10.48550/arXiv.2601.11998.

[28] M. Akram, G. A. Ali, A. Sulaiman, and M. ul Hassan, ‘Accessibility evaluation of Arabic University websites for compliance with success criteria of WCAG 1.0 and WCAG 2.0’, *Univers. Access Inf. Soc.*, vol. 22, no. 4, pp. 1199–1214, Nov. 2023, doi: 10.1007/s10209-022-00921-8.

[29] H. M. Adil, S. Ali, M. Sultan, M. Ashiq, and M. Rafiq, ‘Open education resources’ benefits and challenges in the academic world: a systematic review’, *Glob. Knowl. Mem. Commun.*, vol. 73, no. 3, pp. 274–291, Jul. 2022, doi: 10.1108/GKMC-02-2022-0049.

[30] A. Miralrio, J. Muñoz-Villota, and C. Camacho-Zuñiga, ‘From flexibility to adaptive learning: a pre-COVID-19 perspective on distance education in Latin America’, *Front. Comput. Sci.*, vol. 6, Mar. 2024, doi: 10.3389/fcomp.2024.1250992.

[31] S. O. Boufaida, A. Benmachiche, A. Bennour, M. Maatallah, M. Derdour, and F. Ghabban, ‘Enhancing MOOC Course Classification with Convolutional Neural Networks

via Lion Algorithm-Based Hyperparameter Tuning’, *SN Comput. Sci.*, 2025, doi: https://doi.org/10.1007/s42979-025-04179-8.

[32] A. Benmachiche, B. Tahar, L. M. Tayeb, and Z. Asma, ‘A dynamic navigation for autonomous mobiles robots’, *Intell. Decis. Technol.*, vol. 10, no. 1, pp. 81–91, Jan. 2016, doi: 10.3233/IDT-150239.

[33] I. Boutabia, A. Benmachiche, A. A. Betouil, C. Chemam, and M. Boutassetta, ‘Analyzing Accuracy-Latency Trade-offs in Deep Learning Models for Embedded Vision Systems’, Mar. 01, 2026, *Social Science Research Network, Rochester, NY*: 6435019. Accessed: Apr. 30, 2026. [Online]. Available: https://papers.ssrn.com/abstract=6435019

[34] D. Abbas, A. Benmachiche, M. Derdour, and B. K. Sedraoui, ‘Toward Trustworthy IoT Systems: Integrating Blockchain and Artificial Intelligence’, May 06, 2025, *Social Science Research Network, Rochester, NY*: 6460959. doi: 10.2139/ssrn.6460959.

[35] B. K. Sedraoui, A. Benmachiche, A. Makhlouf, D. Abbas, and M. Derdour, ‘Cybersecurity in E-Learning: A Literature Review on Phishing Detection Using ML and DL Techniques’, in *2025 International Conference on Networking and Advanced Systems (ICNAS)*, Oct. 2025, pp. 1–10. doi: 10.1109/ICNAS68168.2025.11298114.

[36] B. K. Sedraoui, A. Benmachiche, and A. Makhlouf, ‘Developing and Evaluating Lightweight Cryptographic Algorithms for Secure Embedded Systems in IoT Devices’, Jan. 06, 2026, *arXiv*: arXiv:2601.02981. doi: 10.48550/arXiv.2601.02981.

[37] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, W. Ye, Y. Zhang, Y. Chang, P. S. Yu, Q. Yang, and X. Xie, ‘A Survey on Evaluation of Large Language Models’, *ACM Trans Intell Syst Technol*, vol. 15, no. 3, p. 39:1-39:45, Mar. 2024, doi: 10.1145/3641289.

[38] S. Yin, C. Fu, S. Zhao, K. Li, X. Sun, T. Xu, and E. Chen, ‘A survey on multimodal large language models’, *Natl. Sci. Rev.*, vol. 11, no. 12, p. nwae403, Dec. 2024, doi: 10.1093/nsr/nwae403.

[39] F. Z. Canal, T. R. Müller, J. C. Matias, G. G. Scotton, A. R. de S. Junior, E. Pozzebon, and A. C. Sobieranski, ‘A survey on facial emotion recognition techniques: A state-of-the-art literature review’, *Inf. Sci.*, vol. 582, pp. 593–617, Jan. 2022, doi: 10.1016/j.ins.2021.10.005.

[40] G. Manel, S. Bourekkache, O. Kazar, and S. Harous, ‘Generative Artificial Intelligence in Education: Advancing Adaptive and Personalized Learning’, *Acta Inform. Pragensia*, vol. 13, pp. 460–489, Aug. 2024, doi: 10.18267/j.aip.235.

[41] M. Mahmood, W. Jaber Al-kubaisy, and B. Al-Khateeb, ‘Multimedia information retrieval using artificial neural network’, *IAES Int. J. Artif. Intell. IJ-AI*, vol. 12, p. 146, Mar. 2023, doi: 10.11591/ijai.v12.i1.pp146-154.

[42] M. Amina, B. Tahar, B. Abdelmadjid, and M. Majda, ‘Apprentissage évolutionnaire d’une chaîne de Markov cachée pour la reconnaissance automatique de la parole audiovisuelle’, *المجلة الجزائرية للعلوم*, vol. 1, no. 2, pp. 65–73, Jun. 2018.

[43] A. Mellouk and A. Benmachiche, ‘A survey on Navigation Systems in Dynamic Environments’, in *Proceedings of the 10th International Conference on Information Systems and Technologies*, in ICIST ’20. New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 1–7. doi: 10.1145/3447568.3448527.

[44] T. Balqis binti Tengku Abd Rashid, J. bin Sallim, and Y. binti Muhamad Noor, ‘A Comparative Analysis on Artificial Intelligence Techniques for Web Phishing Classification’, *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 769, no. 1, p. 012073, Feb. 2020, doi: 10.1088/1757-899X/769/1/012073.

[45] P. K. Balasamy and K. Athiyappagounder, ‘An Optimized Feature Selection Method for E-Learning Recommender System Using Deep Neural Network based on Multilayer Perceptron.’, *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, 2022, Accessed: Jul. 20, 2024. [Online]. Available: https://inass.org/wp-content/uploads/2022/03/2022103140-2.pdf

[46] B. K. Sedraoui, A. Benmachiche, A. Makhlouf, K. Rais, and C. Chemam, ‘CNN-OOA-Based Cyber Threat Detection: Protecting E-Learning from Phishing’, *Arab. J. Sci. Eng.*, Apr. 2026, doi: 10.1007/s13369-026-11122-3.

[47] I. Boutabia, A. Benmachiche, A. A. Betouil, and C. Chemam, *A Survey in the Use of Deep Learning Techniques in The Open Classroom Approach*. 2024, p. 7. doi: 10.1109/PAIS62114.2024.10541268.

[48] M. Maatallah, A. Benmachiche, K. Rais, and S. Touam, ‘Intelligent Fusion of Multi-Modal Medical Imaging: A Comprehensive Review of Methods, Challenges, and Clinical Integration’, *J. Electron. Electromed. Eng. Med. Inform.*, vol. 8, no. 3, pp. 897–934, May 2026, doi: 10.35882/jeeemi.v8i3.1527.

[49] S. O. Boufaida, A. Benmachiche, and M. Maatallah, ‘Real-Time Image Processing Algorithms for Embedded Systems’, Jan. 09, 2026, *arXiv*: arXiv:2601.06243. doi: 10.48550/arXiv.2601.06243.

[50] X. Chen, X. Wang, Y. Wang, D. Liu, and W. Zhang, ‘Leveraging deep learning and graph analysis for enhanced course recommendations in online education’, *Sci. Rep.*, vol. 15, p. 18623, May 2025, doi: 10.1038/s41598-025-02156-y.

[51] M. Parola, A. Nannini, and S. Poleggi, *Web image search engine based on LSH index and CNN Resnet50*. 2021. doi: 10.48550/arXiv.2108.13301.

[52] S. Zitan, I. Zeroual, and S. Agoujil, ‘Performance Investigation of a Proposed CBIR Search Engine Using Deep Convolutional Neural Networks’, in *Business Intelligence*, Cham: Springer International Publishing, 2022, pp. 41–49. doi: 10.1007/978-3-031-06458-6_3.

[53] M. Jangid and R. Kumar, ‘Deep learning approaches to address cold start and long tail challenges in recommendation systems: a systematic review’, *Multimed. Tools Appl.*, vol. 84, no. 5, pp. 2293–2325, Feb. 2025, doi: 10.1007/s11042-024-20262-3.

[54] R. Sarasu, T. K K, and N. Shanker, ‘SF-CNN: Deep Text Classification and Retrieval for Text Documents’, *Intell. Autom. Soft Comput.*, vol. 35, pp. 1799–1813, Jan. 2023, doi: 10.32604/iasc.2023.027429.

[55] S. O. Boufaida, A. Benmachiche, M. Maatallah, and C. Chemam, ‘An Extensive Examination of Varied Approaches in E-Learning and MOOC Research: A Thorough Overview’, in *2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, IEEE, 2024, pp. 1–8. doi: https://doi.org/10.1109/PAIS62114.2024.10541129.

[56] L. Chen, P. Chen, and Z. Lin, ‘Artificial Intelligence in Education: A Review’, *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.

[57] A. Alam, ‘Employing Adaptive Learning and Intelligent Tutoring Robots for Virtual Classrooms and Smart Campuses: Reforming Education in the Age of Artificial Intelligence’, in *Advanced Computing and Intelligent Technologies*, Singapore: Springer Nature, 2022, pp. 395–406. doi: 10.1007/978-981-19-2980-9_32.

[58] J. Vassoyan, J.-J. Vie, and P. Lemberger, ‘Towards Scalable Adaptive Learning with Graph Neural Networks and Reinforcement Learning’, May 10, 2023, *arXiv*: arXiv:2305.06398. doi: 10.48550/arXiv.2305.06398.

[59] I. Boutabia, A. Benmachiche, A. A. Betouil, C. Chemam, and M. Boutassetta, ‘Learning-Based Dynamic Obstacle Avoidance: A Reinforcement Learning Framework for Autonomous Mobile Robots’, Mar. 16, 2026, *Social Science Research Network, Rochester, NY*: 6424078. doi: 10.2139/ssrn.6424078.

[60] S. Huang, H. Yang, Y. Yao, X. Lin, and Y. Tu, ‘Deep Adaptive Interest Network: Personalized Recommendation with Context-Aware Learning’, Dec. 24, 2024, *arXiv*: arXiv:2409.02425. doi: 10.48550/arXiv.2409.02425.

[61] T. Kumari, Y. Vardan, P. Giridhar Shambharkar, and Y. Gandhi, ‘Comparative Study on Handwritten Digit Recognition Classifier Using CNN and Machine Learning Algorithms’, in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, Mar. 2022, pp. 882–888. doi: 10.1109/ICCMC53470.2022.9753756.

[62] B. K. Sedraoui, A. Benmachiche, A. Bennour, A. Makhlouf, M. Derdour, and F. Ghabban, ‘LSTM-SWAP: A Hybrid Deep Learning Model for Cheating Detection’, *SN Comput. Sci.*, vol. 6, no. 7, p. 798, Sep. 2025, doi: 10.1007/s42979-025-04334-1.

[63] A. Benmachiche, A. A. Betouil, I. Boutabia, A. Nouari, K. Boumahni, and H. Bouzata, ‘A Fuzzy Navigation Approach Using the Intelligent Lights Algorithm for an Autonomous Mobile Robot’, in *12th International Conference on Information Systems and Advanced Technologies “ICISAT 2022”*, Cham: Springer International Publishing, 2023, pp. 112–121. doi: 10.1007/978-3-031-25344-7_11.

[64] A. Benmachiche, A. Makhlouf, and T. Bouhadada, ‘Optimization learning of hidden Markov model using the bacterial foraging optimization algorithm for speech recognition’, *Int. J. Knowl.-Based Intell. Eng. Syst.*, vol. 24, no. 3, pp. 171–181, Aug. 2020, doi: 10.3233/KES-200039.

[65] B. Li, G. Li, J. Xu, X. Li, X. Liu, M. Wang, and J. Lv, ‘A personalized recommendation framework based on MOOC system integrating deep learning and big data’, *Comput Electr Eng*, vol. 106, no. C, Mar. 2023, doi: 10.1016/j.compeleceng.2022.108571.

[66] J. Baek, N. Chandrasekaran, S. Cucerzan, A. Herring, and S. K. Jauhar, ‘Knowledge-Augmented Large Language Models for Personalized Contextual Query Suggestion’, in *Proceedings of the ACM Web Conference 2024*, in WWW ’24. New York, NY, USA: Association for Computing Machinery, mai 2024, pp. 3355–3366. doi: 10.1145/3589334.3645404.

[67] Z. Zhao, W. Fan, J. Li, Y. Liu, X. Mei, and Y. Wang, ‘Recommender Systems in the Era of Large Language Models (LLMs)’, *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 11, pp. 6889–6907, Nov. 2024, doi: 10.1109/TKDE.2024.3392335.

[68] B. Sangamithra, ‘An Improved Information Retrieval System using Hybrid RNN LSTM for Multiple Search Engines’, *Commun. Appl. Nonlinear Anal.*, vol. 31, no. 5s, pp. 167–180, Jul. 2024, doi: 10.52783/cana.v31.1011.

[69] E. Purificato, L. Boratto, and E. W. D. Luca, ‘User Modeling and User Profiling: A Comprehensive Survey’, Feb. 20, 2024, *arXiv*: arXiv:2402.09660. doi: 10.48550/arXiv.2402.09660.

[70] I. Boutabia, ‘A Trajectory Planning using the Particle Swarm Optimization for an Autonomous Mobile Robot in an Environment with Obstacles’, université chadli ben djedid eltarf, 2022. Accessed: Mar. 30, 2026. [Online]. Available: http://depotucbet.univ-eltarf.dz:4000/handle/123456789/2503

[71] A. Benmachiche, M. Derdour, M. S. Kahil, M. C. Ghanem, and M. Deriche, ‘Adaptive Hybrid PSO–APF Algorithm for Advanced Path Planning in Next-Generation Autonomous Robots’, *Sensors*, vol. 25, no. 18, p. 5742, Jan. 2025, doi: 10.3390/s25185742.

[72] E. W. Haaskjold, ‘Leveraging Artificial Intelligence for Analyzing and Classifying Contractual Project Changes in the Oil and Gas Industry: A Case Study Using Algorithms, Machine Learning and Natural Language Processing’, Master thesis, NTNU, 2024. Accessed: Aug. 11, 2025. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3155551

[73] R. C., ‘A Novel PSO Methodology for Web Documents Retrieval’, *Int. J. Comput. Linguist. Res.*, vol. 10, p. 67, Sep. 2019, doi: 10.6025/jcl/2019/10/3/67-75.

[74] R. C and S. K. S, ‘A PSO Strategy of Finding Relevant Web Documents using a New Similarity Measure’, Mar. 26, 2021, *arXiv*: arXiv:2103.14371. doi: 10.48550/arXiv.2103.14371.

[75] A. Bouamrane, M. Derdour, M. S. Kahil, A. Benmachiche, M. Deriche, and M. Nasor, ‘A Novel Sequential Mobility-Based Federated Learning Framework with Agentic Explainable AI for Histopathology-Based Lung Diagnosis’, in *2026 IEEE 23rd International Multi-Conference on Systems, Signals & Devices (SSD)*, Mar. 2026, pp. 855–864. doi: 10.1109/SSD69655.2026.11558924.

[76] T. A. Baha, M. E. Hajji, Y. Es-saady, and H. Fadili, ‘Towards highly adaptive Edu-Chatbot’, *Procedia Comput. Sci.*, vol. 198, pp. 397–403, Jan. 2022, doi: 10.1016/j.procs.2021.12.260.

[77] A. Mathur, S. Pradhan, P. Soni, D. Patel, and R. Regunathan, ‘Automated Test Case Generation Using T5 and GPT-3’, in *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2023, pp. 1986–1992. doi: 10.1109/ICACCS57279.2023.10112971.

[78] J. J. Bird, A. Ekárt, and D. R. Faria, ‘Chatbot Interaction with Artificial Intelligence: human data augmentation with T5 and language transformer ensemble for text classification’, *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 4, pp. 3129–3144, Apr. 2023, doi: 10.1007/s12652-021-03439-8.

[79] R. I. Sallove, T. Mahalakshmi, V. R. Sre, S. Santhi, and V. Anitha, ‘Vec2T5: A Chatbot Integrating Word2Vec and T5 for Improved Performance’, in *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, Dec. 2024, pp. 1057–1063. doi: 10.1109/ICICNIS64247.2024.10823146.

[80] N. Wang and R. R. A. Issa, ‘Transfer Learning-Based Question Generation for Building a Construction Safety Chatbot’, pp. 688–694, Jan. 2024, doi: 10.1061/9780784485231.082.

[81] S. D, K. S. Thrisha Vaishnavi, S. Hariprakassh, B. Abiram, K. Saranya, and P. Tanwar, ‘An Abstractive Summarization and Conversation Bot using T5 and its Variants’, in *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, Nov. 2023, pp. 432–437. doi: 10.1109/ICAICCIT60255.2023.10465740.

[82] J. Yin, Y. Zhu, T.-T. Goh, W. Wu, and Y. Hu, ‘Using Educational Chatbots with Metacognitive Feedback to Improve Science Learning’, *Appl. Sci.*, vol. 14, no. 20, p. 9345, Jan. 2024, doi: 10.3390/app14209345.

[83] M. Allen, U. Naeem, and S. S. Gill, ‘Q-Module-Bot: A Generative AI-Based Question and Answer Bot for Module Teaching Support’, *IEEE Trans. Educ.*, vol. 67, no. 5, pp. 793–802, Oct. 2024, doi: 10.1109/TE.2024.3435427.

[84] Y. Dan, Z. Lei, Y. Gu, Y. Li, J. Yin, J. Lin, L. Ye, Z. Tie, Y. Zhou, Y. Wang, A. Zhou, Z. Zhou, Q. Chen, J. Zhou, L. He, and X. Qiu, ‘EduChat: A Large-Scale Language Model-based Chatbot System for Intelligent Education’, Aug. 05, 2023, *arXiv*: arXiv:2308.02773. doi: 10.48550/arXiv.2308.02773.

[85] C. Antico, S. Giordano, C. Koyuturk, and D. Ognibene, ‘Unimib Assistant: designing a student-friendly RAG-based chatbot for all their needs’, Nov. 29, 2024, *arXiv*: arXiv:2411.19554. doi: 10.48550/arXiv.2411.19554.

[86] K. Taneja, P. Maiti, S. Kakar, P. Guruprasad, S. Rao, and A. K. Goel, ‘Jill Watson: A Virtual Teaching Assistant powered by ChatGPT’, May 17, 2024, *arXiv*: arXiv:2405.11070. doi: 10.48550/arXiv.2405.11070.

[87] S. Kumar, D. Paikar, K. Sai Vutukuri, H. Ali, S. R. Ainala, A. M. Krishnan, and Y. Zhang, ‘KatzBot: Revolutionizing Academic Chatbot for Enhanced Communication’, Oct. 21, 2024, *arXiv*: arXiv:2410.16385. doi: 10.48550/arXiv.2410.16385.

[88] S. Maity and A. Deroy, ‘Generative AI and Its Impact on Personalized Intelligent Tutoring Systems’, Oct. 14, 2024, *arXiv*: arXiv:2410.10650. doi: 10.48550/arXiv.2410.10650.

[89] L. Liu and V. G. Duffy, ‘Exploring the Future Development of Artificial Intelligence (AI) Applications in Chatbots: A Bibliometric Analysis’, *Int. J. Soc. Robot.*, vol. 15, no. 5, pp. 703–716, May 2023, doi: 10.1007/s12369-022-00956-0.

[90] A. A. Ali, M. Mashhour, A. S. Salama, R. Shoitan, and H. Shaban, ‘Development of an Intelligent Personal Assistant System Based on IoT for People with Disabilities’, *Sustainability*, vol. 15, no. 6, Art. no. 6, Jan. 2023, doi: 10.3390/su15065166.

[91] A. Lekova, P. Tsvetkova, T. Tanev, P. Mitrouchev, and S. Kostova, ‘Making humanoid robots teaching assistants by using natural language processing (NLP) cloud-based services’, *J. Mechatron. Artif. Intell. Eng.*, vol. 3, no. 1, Art. no. 1, Jun. 2022, doi: 10.21595/jmai.2022.22720.

[92] R. D. B, N. S, S. S, and S. U, ‘Language Learning Tool with Speech Recognition Assistance’, in *2025 International Conference on Computing and Communication Technologies (ICCCT)*, Apr. 2025, pp. 1–6. doi: 10.1109/ICCCT63501.2025.11019429.

[93] E. Latif, R. Parasuraman, and X. Zhai, ‘PhysicsAssistant: An LLM-Powered Interactive Learning Robot for Physics Lab Investigations’, Jun. 04, 2024, *arXiv*: arXiv:2403.18721. doi: 10.48550/arXiv.2403.18721.

[94] X. Hu, S. Xu, R. Tong, and A. Graesser, ‘Generative AI in Education: From Foundational Insights to the Socratic Playground for Learning’, Jan. 12, 2025, *arXiv*: arXiv:2501.06682. doi: 10.48550/arXiv.2501.06682.

[95] B. Maheswari and R. Nagarajan, ‘EduChatbot: Implementing educational Chatbot for assisting the teaching-learning process by NLP-based hybrid heuristic adopted deep learning framework’, *Kybernetes*, Jul. 2024, doi: 10.1108/K-01-2024-0103.

[96] A. Makhlouf, A. Benmachiche, and I. Boutabia, ‘Enhanced Autonomous Mobile Robot Navigation Using a Hybrid BFO/PSO Algorithm for Dynamic Obstacle Avoidance’, *Informatica*, vol. 48, no. 17, Art. no. 17, Nov. 2024, doi: 10.31449/inf.v48i17.6716.

[97] D. Cortes, B. Bermejo, and C. Juiz, ‘The use of CNNs in VR/AR/MR/XR: a systematic literature review’, *Virtual Real.*, vol. 28, no. 3, p. 154, Aug. 2024, doi: 10.1007/s10055-024-01044-6.

[98] J. J. Sophia and T. P. Jacob, ‘EDUBOT-A Chatbot For Education in Covid-19 Pandemic and VQAbot Comparison’, in *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Aug. 2021, pp. 1707–1714. doi: 10.1109/ICESC51422.2021.9532611.

[99] E. K. Zadeh and M. Alaeifard, ‘Adaptive Virtual Assistant Interaction through Real-Time Speech Emotion Analysis Using Hybrid Deep Learning Models and Contextual Awareness’, *Int. J. Adv. Hum. Comput. Interact.*, vol. 1, no. 1, Art. no. 1, Jul. 2023.

[100] S. Tjayadi and V. C. Mawardi, ‘Laptop Recommendation Intelligent Virtual Assistant using Recurrent Neural Network with RPA for Data Scraping’, in *2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, Nov. 2022, pp. 1–6. doi: 10.1109/ICITDA55840.2022.9971263.

[101] S. R. Burri, V. V. Ghorpade, V. Dutt, and K. Lipi, ‘The Rise of Virtual Health Assistants: Chatbot-Based Healthcare Support and Counseling Using Recurrent Neural Networks (RNNs)’, in *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Nov. 2023, pp. 811–816. doi: 10.1109/ICTACS59847.2023.10390207.

[102] S. Gasmi, T. Bouhadada, and A. Benmachiche, ‘Survey on Recommendation Systems’, in *Proceedings of the 10th International Conference on Information Systems and Technologies*, in ICIST ’20. New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 1–7. doi: 10.1145/3447568.3448518.

[103] A. Apriyanto, R. Sahirin, and S. Bradford, ‘Implementation of Deep Learning in a Voice Recognition System for Virtual Assistants’, *J. Comput. Sci. Adv.*, vol. 2, no. 6, Art. no. 6, Dec. 2024, doi: 10.70177/jsca.v2i6.1533.

[104] A. Benmachiche, B. Hadjar, I. Boutabia, A. A. Betouil, M. Maatallah, and A. Makhlouf, ‘Development of a biometric authentication platform using voice recognition’, pp. 1–7, Oct. 2022, doi: 10.1109/pais56586.2022.9946890.

[105] S. Naik, Y. Pattewar, R. Mahajan, J. Chauhan, and V. Patil, ‘Voice Assistant for Hospitality Management’, *Int. J. Comput. Sci. Program. Lang.*, vol. 7, no. 1, pp. 20–26, Jul. 2021, doi: 10.37628/ijocspl.v7i1.700.

[106] A. Ponmalar, K. Kausalya, G. Bhavana, S. Aarthi, D. Gokulakrishnan, and A. A. Jose, ‘Implementing Virtual Personal Assistant Through Artificial Intelligence Requirements’, in *2022 International Conference on Computer, Power and Communications (ICCPC)*, Dec. 2022, pp. 552–556. doi: 10.1109/ICCPC55978.2022.10072159.

[107] G. Reddy, D. Inupakutika, R. Mundlamuri, S. Kaghyan, and D. Akopian, ‘Hybrid Machine Learning Approach for Task-Oriented Dialog Systems’, *Int. J. Comput. Appl.*, vol. 186, pp. 35–42, May 2024, doi: 10.5120/ijca2024923679.

[108] B. Liu, M. Li, Z. Ji, H. Li, and J. Luo, ‘Intelligent productivity transformation: corporate market demand forecasting with the aid of an AI virtual assistant’, *J. Organ. End User Comput. JOEUC*, vol. 36, no. 1, pp. 1–27, 2024.

[109] M. Bustos-López, N. Cruz-Ramírez, A. Guerra-Hernández, L. N. Sánchez-Morales, N. A. Cruz-Ramos, and G. Alor-Hernández, ‘Wearables for Engagement Detection

in Learning Environments: A Review', *Biosensors*, vol. 12, no. 7, Art. no. 7, Jul. 2022, doi: 10.3390/bios12070509.

[110] Z. Zhu, X. Zheng, T. Ke, and G. Chai, 'Emotion Recognition in Learning Scenes Supported by Smart Classroom and Its Application', *Trait. Signal*, vol. 40, no. 2, pp. 751–758, Apr. 2023, doi: 10.18280/ts.400235.

[111] A. Naim, 'E-Learning Engagement through Convolution Neural Networks in Business Education', *Eur. J. Innov. Nonform. Educ.*, vol. 2, no. 2, Art. no. 2, Feb. 2022.

[112] M. Dang, G. Liu, H. Li, Q. Xu, X. Wang, and R. Pan, 'Multi-object behaviour recognition based on object detection cascaded image classification in classroom scenes', *Appl. Intell.*, vol. 54, no. 6, pp. 4935–4951, Mar. 2024, doi: 10.1007/s10489-024-05409-x.

[113] P. X. Lam, P. Q. H. Mai, Q. H. Nguyen, T. Pham, T. H. H. Nguyen, and T. H. Nguyen, 'Enhancing educational evaluation through predictive student assessment modeling', *Comput. Educ. Artif. Intell.*, vol. 6, p. 100244, 2024.

[114] M. Bond, H. Khosravi, M. De Laat, N. Bergdahl, V. Negrea, E. Oxley, P. Pham, S. Wang Chong and G. Siemens, 'A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour', *Int. J. Educ. Technol. High. Educ.*, vol. 21, no. 1, p. 4, Jan. 2024, doi: 10.1186/s41239-023-00436-z.

[115] K. He and K. Gao, 'Analysis of Concentration in English Education Learning Based on CNN Model', *Sci. Program.*, vol. 2022, p. e1489832, Jul. 2022, doi: 10.1155/2022/1489832.

[116] A. Apicella, P. Arpaia, M. Frosolone, G. Improta, N. Moccaldi, and A. Pollastro, 'EEG-based measurement system for monitoring student engagement in learning 4.0', *Sci. Rep.*, vol. 12, no. 1, p. 5857, Apr. 2022, doi: 10.1038/s41598-022-09578-y.

[117] M. Carroll, M. Ruble, M. Dranias, S. Rebensky, M. Chaparro, J. Chiang, and B. Winslow, 'Automatic Detection of Learner Engagement Using Machine Learning and Wearable Sensors', *J. Behav. Brain Sci.*, vol. 10, no. 3, Art. no. 3, Mar. 2020, doi: 10.4236/jbbs.2020.103010.

[118] S. Zhang, Y. Yang, C. Chen, X. Zhang, Q. Leng, and X. Zhao, 'Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic

review of recent advancements and future prospects', *Expert Syst. Appl.*, vol. 237, p. 121692, Mar. 2024, doi: 10.1016/j.eswa.2023.121692.

[119] S. Ananthan, N. Gao, and F. D. Salim, 'Understanding Physiological Responses of Students Over Different Courses', in *Proceedings of the 2024 ACM International Symposium on Wearable Computers*, Oct. 2024, pp. 104–110. doi: 10.1145/3675095.3676620.

[120] A. Singh, N. Verma, K. Goyal, A. Singh, P. Kumar, and X. Li, 'VisioPhysioENet: Multimodal Engagement Detection using Visual and Physiological Signals', Feb. 03, 2025, *arXiv*: arXiv:2409.16126. doi: 10.48550/arXiv.2409.16126.

[121] S. O. Boufaida, A. Benmachiche, M. Maatallah, and C. Chemam, 'Hybrid Multi-Factor Authentication (MFA) Using Biometrics and Behavioral Analysis', Feb. 23, 2026, *Social Science Research Network, Rochester, NY*: 6295298. doi: 10.2139/ssrn.6295298.

[122] M. S. Fahad, S. Singh, Abhinav, A. Ranjan, and A. Deepak, 'Emotion recognition from spontaneous speech using emotional vowel-like regions', *Multimed. Tools Appl.*, vol. 81, no. 10, pp. 14025–14043, Apr. 2022, doi: 10.1007/s11042-022-12453-7.

[123] G. Fenu, R. Galici, M. Marras, and D. Reforgiato, 'Exploring Student Interactions with AI in Programming Training', in *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, in UMAP Adjunct '24. New York, NY, USA: Association for Computing Machinery, juin 2024, pp. 555–560. doi: 10.1145/3631700.3665227.

[124] R. Schiller, J. Fleckenstein, U. Mertens, A. Horbach, and J. Meyer, 'Understanding the effectiveness of automated feedback: Using process data to uncover the role of behavioral engagement', *Comput. Educ.*, vol. 223, p. 105163, Dec. 2024, doi: 10.1016/j.compedu.2024.105163.

[125] C. Pabba and P. Kumar, 'A vision-based multi-cues approach for individual students' and overall class engagement monitoring in smart classroom environments', *Multimed. Tools Appl.*, vol. 83, no. 17, pp. 52621–52652, May 2024, doi: 10.1007/s11042-023-17533-w.

[126] M. Mosia, F. O. Egara, F. Nannim, and M. Basitere, 'Bayesian Growth Curve Modelling of Student Academic Trajectories: The Impact of Individual-Level Characteristics

and Implications for Education Policy', *Appl. Sci.*, vol. 15, no. 3, Art. no. 3, Jan. 2025, doi: 10.3390/app15031426.

[127] R. Bertolini, S. J. Finch, and R. H. Nehm, 'An application of Bayesian inference to examine student retention and attrition in the STEM classroom', *Front. Educ.*, vol. 8, Feb. 2023, doi: 10.3389/feduc.2023.1073829.

[128] U. Maier and C. Klotz, 'Personalized feedback in digital learning environments: Classification framework and literature review', *Comput. Educ. Artif. Intell.*, vol. 3, p. 100080, Jan. 2022, doi: 10.1016/j.caeai.2022.100080.

[129] Y. Sun, 'Hierarchical Bayesian Knowledge Tracing in Undergraduate Engineering Education', May 29, 2025, *arXiv*: arXiv:2506.00057. doi: 10.48550/arXiv.2506.00057.

[130] M. Mosia, F. Egara, N. Fadip, and M. Basitere, 'Bayesian Hierarchical Modelling of Student Academic Performance: The Impact of Mathematics Competency, Institutional Context, and Temporal Variability', *Educ. Sci.*, vol. 15, pp. 1–18, Feb. 2025, doi: 10.3390/educsci15020177.

[131] M. Mosia and F. Egara, 'PREDICTORS OF STUDENT SUCCESS IN MATHEMATICS: HIERARCHICAL BAYESIAN APPROACH', *Int. J. Eng. Appl. Sci. Technol.*, vol. 6, pp. 59–72, Jun. 2024.

[132] V. Hegde, N. Surendran, and V. M, 'Predicting Student Failure using Peer-based Evaluation and Ratings', in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–6. doi: 10.1109/ICCCNT56998.2023.10308388.

[133] M. A. Reza, 'Educational Outcomes Through Predictive Modeling of Student Performance', Mar. 23, 2025, *Social Science Research Network, Rochester, NY*: 5190301. doi: 10.2139/ssrn.5190301.

[134] Y. Chen, J. Sun, J. Wang, L. Zhao, X. Song, and L. Zhai, 'Machine Learning-Driven Student Performance Prediction for Enhancing Tiered Instruction', Feb. 05, 2025, *arXiv*: arXiv:2502.03143. doi: 10.48550/arXiv.2502.03143.

[135] N. A. Butt, Z. Mehmood, K. Shakeel, S. Alfarhood, M. Safran, and I. Ashraf, 'Performance Prediction of Students in Higher Education Using Multi-Model Ensemble Approach', *IEEE Access*, vol. PP, pp. 1–1, Jan. 2023, doi: 10.1109/ACCESS.2023.3336987.

[136] B. Abdelmdjid, B. Tahar, M. Amina, and M. Majda, 'Un modèle de planification multi robots base sur les algorithmes géné-tiques pour la navigation optimale de robots autonomes', *المجلة الجزائرية للعلوم*, vol. 5, no. 10, pp. 47–50, Jun. 2020.

[137] J. Wang and Y. Yu, 'Machine learning approach to student performance prediction of online learning', *PLOS ONE*, vol. 20, no. 1, p. e0299018, Jan. 2025, doi: 10.1371/journal.pone.0299018.

[138] M. Ilić, V. Mikić, L. Kopanja, and B. Vesin, 'Intelligent techniques in e-learning: a literature review', *Artif. Intell. Rev.*, vol. 56, no. 12, pp. 14907–14953, Dec. 2023, doi: 10.1007/s10462-023-10508-1.

[139] S. Wang, F. Wang, Z. Zhu, J. Wang, T. Tran, and Z. Du, 'Artificial intelligence in education: A systematic literature review', *Expert Syst. Appl.*, vol. 252, p. 124167, Oct. 2024, doi: 10.1016/j.eswa.2024.124167.

[140] T. Cardona, E. A. Cudney, R. Hoerl, and J. Snyder, 'Data Mining and Machine Learning Retention Models in Higher Education', *J. Coll. Stud. Retent.*, vol. 25, no. 1, pp. 51–75, May 2023, doi: 10.1177/1521025120964920.

[141] T. Peng, Y. Liang, W. Wu, J. Ren, Z. Pengrui, and Y. Pu, 'CLGT: A Graph Transformer for Student Performance Prediction in Collaborative Learning', Jul. 30, 2023, *arXiv*: arXiv:2308.02038. doi: 10.48550/arXiv.2308.02038.

[142] A. E. Hajoui, O. Y. Alaoui, O. E. Kharki, M. Wahbi, H. Boulassal, and M. Maatouk, 'An Early Student Performance Prediction Using GRU Model', in *2024 International Conference on Circuit, Systems and Communication (ICCSC)*, Jun. 2024, pp. 1–6. doi: 10.1109/ICCSC62074.2024.10617020.

[143] M. Delianidi, K. Diamantaras, G. Chrysogonidis, and V. Nikiforidis, 'Student Performance Prediction Using Dynamic Neural Models', arXiv.org. Accessed: Mar. 31, 2025. [Online]. Available: https://arxiv.org/abs/2106.00524v1

[144] D. Babik, E. Gehringer, J. Kidd, K. Sunday, D. Tinapple, and S. Gilbert, 'A systematic review of educational online peer-review and assessment systems: charting the

landscape', *Educ. Technol. Res. Dev.*, vol. 72, no. 3, pp. 1653–1689, Jun. 2024, doi: 10.1007/s11423-024-10349-x.

[145] M. W. Gazica, G. D. Leto, and A. L. Irish, 'The effects of unexpected changes to content delivery on student learning outcomes: A psychological contract perspective during the COVID-19 era', *Psychol. Sch.*, vol. 59, no. 8, pp. 1473–1491, 2022, doi: 10.1002/pits.22685.

[146] M. Shoaib, N. Sayed, J. Singh, J. Shafi, S. Khan, and F. Ali, 'AI student success predictor: Enhancing personalized learning in campus management systems', *Comput. Hum. Behav.*, vol. 158, p. 108301, Sep. 2024, doi: 10.1016/j.chb.2024.108301.

[147] K. Alalawi, R. Athauda, R. Chiong, and I. Renner, 'Evaluating the student performance prediction and action framework through a learning analytics intervention study', *Educ. Inf. Technol.*, vol. 30, no. 3, pp. 2887–2916, Feb. 2025, doi: 10.1007/s10639-024-12923-5.

[148] A. Kukkar, R. Mohana, A. Sharma, and A. Nayyar, 'A novel methodology using RNN + LSTM + ML for predicting student's academic performance', *Educ. Inf. Technol.*, vol. 29, no. 11, pp. 14365–14401, Aug. 2024, doi: 10.1007/s10639-023-12394-0.

[149] F. Ouyang, M. Wu, L. Zheng, L. Zhang, and P. Jiao, 'Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course', *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, p. 4, Jan. 2023, doi: 10.1186/s41239-022-00372-4.

[150] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, 'Predicting academic performance of students from VLE big data using deep learning models', *Comput. Hum. Behav.*, vol. 104, p. 106189, Mar. 2020, doi: 10.1016/j.chb.2019.106189.

[151] I. Boutabia, A. Benmachiche, A. A. Betouil, C. Chemam, and K. Rais, 'Advanced Text Prediction System Integrated Within the Search Engine for the Open Classroom Approach Based on Particle Swarm Optimization and Long Short-Term Memory Models', *Arab. J. Sci. Eng.*, Mar. 2026, doi: 10.1007/s13369-026-11247-5.

[152] 'Welcome to Python.org', Python.org. Accessed: Aug. 14, 2025. [Online]. Available: https://www.python.org/

[153] 'TensorFlow', TensorFlow. Accessed: Aug. 14, 2025. [Online]. Available: https://www.tensorflow.org/?hl=fr

[154] *pyswarms: A Python-based Particle Swarm Optimization (PSO) library.* Python. Accessed: Aug. 14, 2025. [Online]. Available: https://github.com/ljvmiranda921/pyswarms

[155] 'NLTK :: Natural Language Toolkit'. Accessed: Aug. 14, 2025. [Online]. Available: https://www.nltk.org/

[156] 'pandas - Python Data Analysis Library'. Accessed: Aug. 14, 2025. [Online]. Available: https://pandas.pydata.org/

[157] 'NumPy'. Accessed: Aug. 14, 2025. [Online]. Available: https://numpy.org/

[158] 'Matplotlib — Visualization with Python'. Accessed: Aug. 14, 2025. [Online]. Available: https://matplotlib.org/

[159] S. R. Sayak Paul, 'arXiv Paper Abstracts'.

[160] 'University Chatbot Dataset'. Accessed: Mar. 12, 2025. [Online]. Available: https://www.kaggle.com/datasets/tusharpaul2001/university-chatbot-dataset

[161] I. Boutabia, A. Benmachiche, A. Bennour, A. A. Betouil, M. Derdour, and F. Ghabban, 'Hybrid CNN-ViT Model for Student Engagement Detection in Open Classroom Environments', *SN Comput. Sci.*, vol. 6, no. 6, p. 684, Jul. 2025, doi: 10.1007/s42979-025-04228-2.

[162] K. Delgado, J. M. Origgi, T. Hasanpoor; H. Yu, D. Allessio, I. Arroyo, W. Lee, M. Betke, B. Woolf, S. Adel Bargal, 'Student Engagement Dataset', in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021, pp. 3621–3629. doi: 10.1109/ICCVW54120.2021.00405.

[163] 'Intelligent Classroom Dataset'. Accessed: Apr. 05, 2025. [Online]. Available: https://www.kaggle.com/datasets/ziya07/intelligent-classroom-dataset

[164] M. Redjimi, A. Makhlouf, M. Maatallah, A. Benmachiche, and S. O. Boufaida, 'Enhancing Malware Detection Robustness through Hybrid Adversarial Machine Learning', Apr. 30, 2025, *Social Science Research Network, Rochester, NY*: 6722779. doi: 10.2139/ssrn.6722779.

[165] D. Abbas, A. Benmachiche, M. Derdour, and B. K. Sedraoui, ‘Privacy and Security in Decentralized Cyber-Physical Systems: A Survey’, in *2025 International Conference on Networking and Advanced Systems (ICNAS)*, Oct. 2025, pp. 1–10. doi: 10.1109/ICNAS68168.2025.11297996.

[166] B. K. Sedraoui, A. Benmachiche, A. Makhlouf, and C. Chemam, ‘Applications of Secure Multi-Party Computation in Financial Services’, Jan. 01, 2026, *arXiv*: arXiv:2601.00334. doi: 10.48550/arXiv.2601.00334.

[167] B. K. Sedraoui, A. Benmachiche, A. Makhlouf, and C. Chemam, ‘Exploring the Integration of Differential Privacy in Cybersecurity Analytics: Balancing Data Utility and Privacy in Threat Intelligence’, Jan. 01, 2026, *arXiv*: arXiv:2601.00385. doi: 10.48550/arXiv.2601.00385.