10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

Harnessing AI for effective language teaching: Innovations in Hindi Urdu Instructions.

Shagufta Fatema

Ph.D, Middle East South Asia Program, University of California, Davis CA, USA

Abstract: The integration of Artificial Intelligence (AI) into language education has opened new horizons for enhancing instructional methodologies, particularly in the context of low-resource and morphologically rich languages such as Hindi and Urdu. This review paper explores the current innovations, applications, and challenges associated with harnessing AI for effective language teaching in Hindi-Urdu instructional settings. It examines various AI-driven tools including Natural Language Processing (NLP) applications, intelligent tutoring systems, adaptive learning platforms, automated assessment tools, and conversational agents like chatbots. The paper highlights how these technologies facilitate personalized learning, real-time feedback, grammar and vocabulary enhancement, and pronunciation improvement, while also catering to diverse learner profiles and regional variations in language usage. Emphasis is placed on the pedagogical implications of AI, such as increased learner autonomy, engagement, and motivation, particularly in e-learning and blended learning environments. Case studies and existing platforms are analyzed to demonstrate the real-world implementation of AI in Hindi-Urdu classrooms, along with limitations like data scarcity, script variation, and dialectal diversity. Furthermore, the review discusses the ethical and sociolinguistic considerations involved in deploying AI tools in language teaching, including issues of bias, digital divide, and teacher-student interaction dynamics. The paper concludes by identifying research gaps and proposing future directions for the development of AI-based language instruction systems that are culturally relevant, linguistically inclusive, and pedagogically effective for Hindi and Urdu learners. This work contributes to the evolving discourse on AI in education by foregrounding the unique challenges and opportunities in the context of South Asian languages.

Keywords: Artificial Intelligence, Language Teaching, Hindi-Urdu Instruction, NLP, Educational Technology, Chatbots, Adaptive Learning, E-Learning, Low-Resource Languages, AI in Education.

1. Introduction

The advent of Artificial Intelligence (AI) [1] has revolutionized numerous domains, and education is no exception. Within the educational landscape, language learning stands out as one of the areas undergoing a profound transformation due to AI-driven innovations. Language acquisition is a complex cognitive and social process that benefits greatly from personalized feedback, contextual understanding, and adaptive content delivery—all of which are strengths of AI technologies. AI-based systems, including natural language processing (NLP) tools, intelligent tutoring systems, and speech recognition engines, are increasingly being employed to enhance the quality and accessibility of language education. These systems not only mimic human interaction but also provide scalable solutions to challenges such as large class sizes, lack of qualified instructors, and the need for individualized learning. The emergence of such tools represents a paradigm shift from traditional pedagogical approaches to more dynamic, interactive, and learner-centric models of instruction [1].

The global popularity of English and major European languages such as French, German, and Spanish has historically attracted disproportionate attention in the realm of language technology development. Consequently, a significant portion of existing AI-powered educational tools is tailored for these high-resource languages [2]. However, as the world moves toward greater linguistic inclusion and equity in digital education, it becomes imperative to extend the benefits of AI to languages that have traditionally been marginalized in technological development. Hindi and Urdu—two linguistically rich and widely spoken South Asian languages—fall into this category of low-resource languages from a technological standpoint. Despite their cultural, literary, and demographic significance, these languages lag behind in terms of computational tools, annotated datasets, and advanced language learning platforms [2].

Hindi and Urdu collectively represent over half a billion speakers worldwide. Hindi, written in the Devanagari script, is the official language of India and is used extensively in education, administration, and media. Urdu,

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

primarily spoken in Pakistan and various parts of India, uses the Perso-Arabic Nastaliq script and holds a significant place in South Asian literature, poetry, and cultural discourse. While both languages share a common linguistic base and are mutually intelligible in spoken form, their distinct scripts and socio-political trajectories have led to divergent development in educational resources and digital tools. The integration of AI into Hindi-Urdu language teaching, therefore, not only demands a nuanced understanding of linguistic structures but also requires sensitivity to socio-cultural, political, and pedagogical contexts [2]. One of the primary motivations for employing AI in Hindi-Urdu instruction is the potential to democratize access to quality language education. In countries like India and Pakistan, educational inequalities are exacerbated by factors such as geographic location, economic disparity, and inconsistent quality of schooling. Rural and underserved areas often lack trained language teachers, and instructional materials are frequently outdated or misaligned with students' needs. AI-powered platforms can help bridge this gap by offering customized learning experiences, interactive language drills, real-time feedback, and 24/7 accessibility. Such systems can also facilitate self-paced learning, which is particularly beneficial for adult learners, working professionals, or individuals pursuing education in non-traditional settings.

Furthermore, AI can address the issue of learner diversity in multilingual classrooms. In South Asia, it is common for students to be exposed to multiple languages from early childhood [3]. In such linguistically plural environments, learners may possess varying levels of proficiency in Hindi and Urdu, often influenced by regional dialects, socio-economic backgrounds, and mother tongues. AI-enabled adaptive learning tools can dynamically assess a learner's current language level and modify instruction accordingly, thus enhancing learner engagement and reducing cognitive overload. Through machine learning algorithms, these platforms can analyze user input, detect errors, and personalize content delivery—ranging from vocabulary exercises to grammar tutorials and pronunciation guides [3].

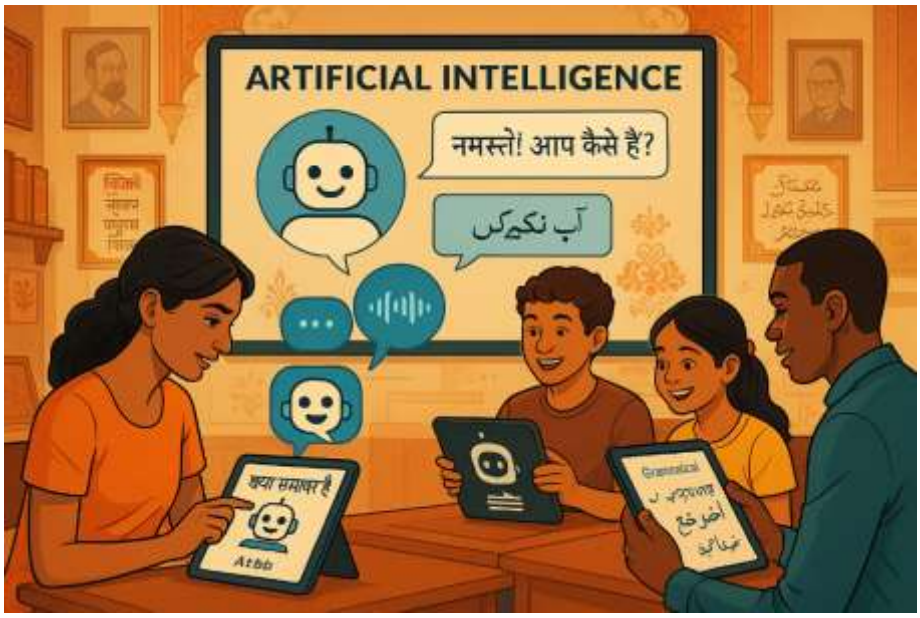

Figure 1. AI-Powered Hindi-Urdu Language Learning in a Multicultural Classroom

Another compelling motivation for integrating AI into Hindi-Urdu language education lies in the urgent need to preserve linguistic and literary heritage. Languages such as Hindi and Urdu are custodians of vast repositories of classical literature, poetry, folklore, and oral traditions. However, globalization and the increasing dominance of English have led to a decline in the intergenerational transmission of these languages, particularly in urban settings. By leveraging AI for language education, it becomes possible not only to teach the functional aspects of the language but also to embed rich cultural content in the learning process. AI-generated content, automated storytelling systems, and interactive literary analysis tools can foster deeper engagement with classical and contemporary Hindi-Urdu literature, thereby revitalizing interest among younger generations [4]. Despite the promises, there are challenges specific to deploying AI in low-resource language instruction. Unlike English or

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

Spanish, for which extensive corpora and pre-trained language models are available, Hindi and Urdu suffer from a paucity of digitized, annotated linguistic resources. This scarcity hinders the training and fine-tuning of AI models, resulting in reduced accuracy and effectiveness of educational tools. Additionally, script processing poses a technical hurdle. Hindi's Devanagari script includes complex ligatures and diacritics, while Urdu's Nastaliq script is cursive, non-linear, and notoriously difficult to render and parse computationally. These challenges necessitate dedicated efforts in developing script-specific OCR (Optical Character Recognition), speech-to-text systems, and NLP pipelines tailored to the unique properties of these languages [4].

Moreover, the sociolinguistic landscape of Hindi and Urdu adds further complexity. Dialectal variations are widespread; for instance, Hindi is spoken in forms such as Awadhi, Bhojpuri, Braj, and Khari Boli, each with its phonetic and lexical peculiarities. Urdu too encompasses dialects like Dakhni and Rekhta, which deviate significantly from standard formal usage. AI models must be trained to recognize and adapt to these variations to ensure inclusivity and effectiveness in teaching. Without this flexibility, AI-based platforms risk alienating learners who do not speak standardized dialects, thus reinforcing linguistic hierarchies and educational inequalities [5]. There is also a pedagogical dimension to the integration of AI in Hindi-Urdu instruction. Language learning is not merely a mechanical acquisition of rules and vocabulary; it involves emotional resonance, social interaction, and contextual understanding. While AI can simulate aspects of human tutoring, it cannot fully replace the nuanced feedback, empathy, and cultural knowledge provided by skilled teachers. Therefore, the goal should not be to replace traditional instruction but to supplement it through AI-assisted tools that enhance engagement, feedback, and learner autonomy. Effective integration requires a hybrid model that leverages the strengths of both human and artificial intelligence [5].

From a policy perspective, there is a growing recognition of the need to promote regional languages and ensure their representation in the digital domain. Initiatives such as India's National Education Policy (NEP) 2020 emphasize mother-tongue instruction and multilingual education, creating a conducive environment for AI interventions in regional language pedagogy. International organizations such as UNESCO have also underscored the importance of digital language inclusion [6]. However, translating policy into practice requires coordinated efforts between governments, academic institutions, technology developers, and local communities. Investment in open-source linguistic datasets, collaborative platform development, and teacher training programs is essential to build sustainable AI ecosystems for Hindi-Urdu instruction. In recent years, some promising initiatives have emerged. Platforms like Google Translate and Microsoft Azure's Cognitive Services have gradually expanded their support for Hindi and Urdu, though often with limited contextual understanding. More specialized applications, such as Hindi and Urdu voice assistants, AI-powered grammar correction tools, and educational apps, are beginning to surface, though they remain in the early stages of development. These tools provide a glimpse into the possibilities of AI in regional language education but also highlight the gaps that need to be addressed in terms of accuracy, usability, and cultural relevance [6].

In conclusion, the use of Artificial Intelligence in language education presents a transformative opportunity for Hindi and Urdu, especially given their status as underrepresented languages in the AI domain. The integration of AI can make language instruction more accessible, personalized, and engaging, while also contributing to the preservation and promotion of linguistic heritage. However, to fully realize this potential, it is necessary to confront the challenges posed by resource limitations, script complexities, and dialectal diversity. This review paper aims to provide a comprehensive overview of existing innovations, ongoing challenges, and future directions in harnessing AI for effective Hindi-Urdu instruction. By doing so, it contributes to the broader discourse on inclusive, equitable, and culturally sensitive AI in education.

1.1 Objectives

The study focuses on the following objectives:

- To explore how Artificial Intelligence (AI) is being used in Hindi and Urdu language teaching.
- To identify innovative AI tools and techniques that support language learning.
- To highlight the unique challenges of teaching Hindi and Urdu using AI.

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

- To review existing platforms and case studies using AI in language instruction.
- To discuss the benefits of AI for personalized, accessible, and engaging learning.
- To suggest future directions for improving AI-based Hindi-Urdu educational tools.

.

2. Literature Review

Chandio, F. et al. (2022) [7]: The authors present a comprehensive systematic literature review on Urdu sentiment analysis using various techniques, including traditional machine learning algorithms, deep learning models (like LSTM and BiLSTM), and transfer learning approaches. A major focus is placed on handling Roman Urdu and the associated challenges such as non-standardized spellings and code-mixed texts. This review also highlights the scarcity of large, annotated sentiment datasets for Urdu and emphasizes the role of preprocessing in improving model performance. Though not solely focused on education, the study offers vital insights into affective computing in low-resource languages, which is essential for developing emotionally aware chatbots and feedback mechanisms in AI-supported Hindi-Urdu learning platforms.

Khatri, S. K., et al. (2022) [8]: This PRISMA-guided review compares speech recognition advancements across English, Mandarin, Hindi, and Urdu. It reveals a significant gap in available speech data, especially for Urdu, which has less than 10 hours of training data, resulting in high error rates. The authors advocate for developing larger datasets and more robust deep learning models tailored to South Asian languages. Their findings are critical for designing AI-powered pronunciation practice tools and oral response assessments in Hindi-Urdu educational applications.

Kashif, M. (2021) [9]: The paper proposes a ResNet-18-based solution for recognizing handwritten Urdu text—a longstanding challenge due to the complex Nastaliq script. Using the UNHD dataset, the model achieved high recognition accuracy, showcasing the strength of deep learning for OCR in cursive scripts. This research is pivotal for integrating handwriting analysis into AI-driven language learning platforms, enabling features such as handwriting practice, digital worksheets, and automated corrections in Urdu instruction.

Khalid, S., et al. (2021) [10]: address the challenge of word similarity in Urdu, a low-resource language, by developing word embeddings using the FastText model. Their research focuses on improving semantic similarity tasks by training models on a larger and more diverse Urdu corpus compiled through web crawling. The authors demonstrate that FastText embeddings significantly outperform skip-gram models when dealing with morphologically rich languages like Urdu. This study is particularly important for AI-based language education because it strengthens the semantic core of NLP systems, which are essential for developing grammar checkers, vocabulary games, and automated content generation in Urdu learning platforms.

Asim, M. N., et al. (2020) [11]: This study benchmarks various machine learning and deep learning approaches for Urdu text classification. The authors construct a novel labeled corpus and compare the performance of models including SVM, Naive Bayes, CNN, and fine-tuned BERT for Urdu. The findings suggest that hybrid methods combining traditional feature engineering with deep learning yield superior results. These insights are especially relevant for AI-powered Hindi-Urdu platforms that adaptively recommend content or evaluate learner responses based on document or phrase classification.

Table 1. Literature Review Findings

Author Name (Year)	Main Concept	Findings	Limitations
Chandio et al. (2022)	Review of Urdu sentiment analysis methods, including ML, DL, and Roman Urdu challenges	LSTM, BiLSTM, and lexicon-based methods are effective; preprocessing plays a crucial role	Lack of large annotated datasets; difficulty in handling Roman Urdu and code-mixing

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

Khatri et al. (2022)	Speech recognition comparison across English, Mandarin, Hindi, and Urdu using PRISMA	Urdu has the least amount of training data, leading to high error rates; highlights need for dataset development	Severe shortage of spontaneous Urdu speech corpora; no implementation of proposed models
Kashif (2021)	Handwritten Urdu text recognition using ResNet-18	High recognition accuracy on UNHD dataset; effective handling of Nastaliq script with CNN	Limited to one dataset; lacks testing across different handwriting styles and real-time use
Khalid et al. (2021)	FastText embeddings for Urdu word similarity tasks	FastText embeddings significantly outperform skip-gram for morphologically rich Urdu	Evaluation limited to word similarity; lacks integration with full NLP pipelines
Asim et al. (2020)	Urdu text classification using ML, DL, and BERT models	Hybrid models combining traditional ML and DL yield the best accuracy	Dataset limited in diversity; lacks task-specific tuning and generalization studies

Despite notable advancements in applying Artificial Intelligence (AI) to Hindi and Urdu language processing, significant research gaps remain that limit the development of effective educational tools. A primary gap lies in the lack of large, diverse, and annotated datasets, especially for Urdu, which hampers the performance of machine learning and deep learning models in real-world applications. Most studies focus on isolated tasks such as sentiment analysis, text classification, or word similarity, without integrating these components into comprehensive language learning systems. Furthermore, limited support for multi-script processing (e.g., Devanagari for Hindi and Nastaliq for Urdu) and dialectal variations restricts the adaptability of current models to varied learner contexts. There is also a shortage of context-aware, interactive platforms—such as chatbots or adaptive tutoring systems—designed specifically for Hindi-Urdu pedagogy. Additionally, Roman Urdu and code-mixed input, which are common in informal and digital communication, are still underexplored in educational contexts. These gaps underscore the urgent need for cross-disciplinary efforts to develop inclusive, scalable, and pedagogically grounded AI tools for Hindi and Urdu language instruction.

3. Overview of Hindi and Urdu Language Instruction

Hindi and Urdu are two of the most prominent languages of South Asia, sharing a common grammatical foundation but differing significantly in script, vocabulary, and socio-cultural identity. Linguistically, both belong to the Indo-Aryan branch of the Indo-European language family and are mutually intelligible in their colloquial spoken forms. Often collectively referred to as "Hindustani" in spoken registers, Hindi and Urdu differ primarily in their written scripts and formal vocabulary. Hindi uses the Devanagari script, which is syllabic and phonetic in nature, while Urdu employs the Perso-Arabic Nastaliq script, known for its ornate and cursive form. Lexically, Hindi borrows heavily from Sanskrit, whereas Urdu draws extensively from Persian and Arabic. Despite their linguistic overlap, the languages are taught and treated separately within educational and governmental systems in India and Pakistan, respectively, and increasingly in diaspora communities [12]. The linguistic richness of Hindi and Urdu presents both opportunities and challenges for instruction. Both languages exhibit complex morphology, including a system of gendered nouns, postpositions, and honorifics. Verb conjugations are sensitive to gender, number, and politeness level, making the process of acquiring fluency non-trivial, particularly for second-language learners. Additionally, sentence structure is flexible and context-dependent, allowing for various syntactic constructions that must be mastered through extensive exposure and practice. Phonologically, the languages include retroflex consonants and aspirated sounds that may not exist in other languages, requiring specialized attention during pronunciation instruction. These features, while making the languages expressive and culturally nuanced, pose distinct challenges for learners and educators alike, particularly when integrating language technologies into the instructional process [13].

Traditionally, the teaching of Hindi and Urdu has relied on teacher-centered, grammar-translation methods that emphasize rote learning and memorization. Instruction often begins with a focus on mastering the alphabet, followed by the recognition and writing of words and sentences. Grammar rules are typically taught in isolation, and learners are expected to apply these rules in written exercises. In both languages, considerable emphasis is placed on the reading and interpretation of literary texts—poems, stories, and classical prose—rather than functional language skills such as speaking, listening, or real-world communication. As a result, students may

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

acquire a theoretical understanding of grammar without developing sufficient proficiency in oral or practical use of the language. In Urdu instruction, this issue is compounded by the challenge of reading and writing in the Nastaliq script, which, due to its intricate calligraphic structure, demands more visual recognition skills and writing practice than Latin-based scripts or even Devanagari [14]. In the context of Urdu education, script complexity often leads to a heavy reliance on oral instruction or Roman Urdu (Urdu written in the Latin script), especially in digital communication and informal settings. However, this shift to Roman Urdu has not been formally integrated into classroom instruction, leading to a disconnect between learners' daily language use and academic expectations. Similarly, in Hindi instruction, while the Devanagari script benefits from better digital support and standardization, students often face challenges in mastering the ligatures and diacritic marks essential to correct reading and pronunciation. Digital tools that support interactive script learning and pronunciation practice are still in developmental stages and are rarely incorporated into mainstream curricula, especially in government or rural schools [15].

An additional complexity in the teaching of Hindi and Urdu arises from the vast dialectal variation present within each language. Hindi is spoken in various forms, including Khari Boli, Braj, Awadhi, Bhojpuri, and Haryanvi, each with distinct phonetic, lexical, and syntactic features. Urdu, too, has regional variants such as Dakhni (spoken in southern India) and Rekhta (a poetic form of Urdu used in classical literature). These dialectal differences often lead to inconsistencies in vocabulary, pronunciation, and grammatical usage, particularly in multilingual classrooms or in regions where Hindi or Urdu is a second language. Most standardized teaching materials do not adequately reflect this linguistic diversity, leading to a mismatch between classroom instruction and students' home language practices. Learners may find formal instruction alienating or confusing when it diverges significantly from the forms of language they use in everyday life. Another pedagogical challenge lies in the lack of integration of modern communicative methods in Hindi and Urdu language instruction. While language pedagogy globally has moved toward more communicative and learner-centered approaches, the teaching of Hindi and Urdu continues to rely heavily on traditional methods with limited student interaction. Speaking and listening skills are rarely emphasized, and classroom activities seldom include role-plays, group discussions, or multimedia resources. This gap is particularly evident in state-run schools and under-resourced institutions, where infrastructural limitations and lack of teacher training further hinder innovation in language pedagogy. In addition, there is often a cultural reluctance to shift away from established teaching traditions, which are perceived as preserving literary and linguistic purity [16].

The situation is further complicated by a general shortage of trained language teachers who are equipped to teach these languages using technology or modern pedagogical tools. Many teachers are not proficient in the use of digital platforms, and teacher training programs rarely include modules on language technology, artificial intelligence, or online teaching strategies. As a result, even when educational institutions have access to digital resources, these tools remain underutilized or are used in ways that do not align with effective language acquisition principles. Moreover, the lack of age-appropriate, curriculum-aligned, and interactive educational content for Hindi and Urdu further limits the integration of technology into classrooms. Most digital tools and language learning apps focus on English and a few other global languages, leaving a significant gap in AI-driven or interactive resources for Hindi and Urdu instruction. In addition to pedagogical and linguistic challenges, there are socio-political factors that influence how Hindi and Urdu are taught and learned. In India, Hindi is promoted as the national language, but this can generate resistance in non-Hindi-speaking states, affecting its adoption in schools outside the Hindi belt. In contrast, Urdu, though constitutionally recognized, often receives limited institutional support, particularly in mainstream Indian education systems. In Pakistan, while Urdu is the national language, English remains the dominant medium of instruction in elite institutions, pushing Urdu into a secondary role. These tensions influence curriculum design, teacher deployment, and the development of educational resources, often leading to imbalanced implementation and underrepresentation in national education policies [17].

In summary, the teaching of Hindi and Urdu is shaped by a complex interplay of linguistic, pedagogical, technological, and socio-political factors. While the languages possess rich grammatical structures and literary traditions, their instruction often relies on outdated methods that fail to engage learners or build practical communication skills. Script complexities, dialectal variations, and a lack of modern teaching resources further exacerbate these issues. The integration of Artificial Intelligence and other technological innovations offers promising solutions, but such efforts must be grounded in a nuanced understanding of these challenges. There is

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

a critical need for the development of inclusive, interactive, and culturally responsive tools that can support the teaching and learning of Hindi and Urdu in diverse contexts.

4. Role of Artificial Intelligence in Language Education

Artificial Intelligence (AI) [18] has emerged as a transformative force in education, offering innovative tools and methodologies that are reshaping traditional language teaching paradigms. The integration of AI into language education brings with it a new realm of possibilities, particularly for low-resource languages like Hindi and Urdu, which have historically been underrepresented in digital education platforms. AI technologies such as Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning have opened doors to personalized learning experiences, real-time feedback, intelligent tutoring systems, and the automation of pedagogical tasks. These tools not only increase efficiency but also provide adaptive, engaging, and data-driven environments that address individual learner needs and overcome long-standing challenges in language instruction.

At the core of AI's contribution to language education is Natural Language Processing, which enables computers to interpret, analyze, generate, and understand human languages. NLP underpins several applications critical to education: automatic speech recognition (ASR), machine translation, grammar checking, sentiment analysis, text summarization, and conversational agents such as chatbots. In the context of Hindi and Urdu, NLP enables the development of applications that can process scripts like Devanagari and Nastaliq, understand syntactic and morphological complexities, and provide contextual feedback to learners. While major progress has been made in NLP for English and a few global languages, recent advancements in transfer learning and multilingual models (like mBERT and XLM-R) are beginning to make sophisticated NLP tools accessible for under-resourced languages, allowing Hindi and Urdu to benefit from cutting-edge technologies with less data-intensive demands [19].

Machine Learning and Deep Learning are the foundational technologies that drive the adaptive and intelligent behavior of AI systems in education. ML algorithms can analyze large datasets of learner behavior, assess patterns in language usage, and make predictions to customize learning paths. In language education, ML powers adaptive testing systems, intelligent error correction, pronunciation analysis, and content recommendation engines. Deep learning, particularly through models like Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Transformers, brings additional depth to AI capabilities by enabling machines to understand linguistic sequences, generate human-like responses, and even evaluate speaking and writing skills with high accuracy. For Hindi and Urdu learners, deep learning models trained on curated corpora can facilitate speech-to-text conversion, automated essay scoring, and real-time translation—all of which enhance the interactivity and accessibility of language instruction [20]. One of the most promising applications of AI in language learning is the intelligent tutoring system (ITS), which mimics the role of a human tutor by offering personalized instruction and feedback. AI-powered ITS platforms can evaluate a learner's strengths and weaknesses, adjust the difficulty level of exercises, and provide timely interventions based on learning behavior. For example, a student struggling with Hindi verb conjugation might be guided through customized drills, visual aids, and contextual examples tailored to their proficiency level. Similarly, a learner of Urdu could receive automatic feedback on script formation, pronunciation accuracy, or grammatical correctness, thus allowing them to progress at their own pace. These systems are particularly valuable in resource-limited settings where qualified language teachers may be scarce or classroom sizes unmanageably large [20].

Another area where AI demonstrates significant pedagogical value is in the creation of conversational agents, or AI chatbots, capable of engaging learners in dialogue practice. Unlike traditional exercises, chatbots offer interactive, low-pressure environments where students can practice conversational Hindi or Urdu, receive feedback on errors, and build confidence in their speaking and listening skills. These agents can be designed to simulate real-life situations, such as buying groceries, attending a job interview, or visiting a doctor, which enhances functional language use. Moreover, by integrating sentiment analysis and emotion recognition, AI can assess learner attitudes and engagement, thereby enabling a more holistic understanding of the learning process.

AI also contributes to the generation and recommendation of educational content. With the help of NLP and ML, systems can automatically generate reading passages, vocabulary quizzes, comprehension tests, and even story-based learning modules suited to the learner's level and interests. This reduces the burden on educators to manually

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

curate or develop content and ensures that learners are exposed to a wider variety of materials. For Hindi and Urdu, where localized and curriculum-aligned digital content is limited, AI can assist in dynamically translating and adapting resources from global repositories into these languages while preserving linguistic nuances.

The pedagogical benefits of AI integration are multifold. Firstly, AI supports personalized learning, which is crucial for language acquisition. Unlike traditional one-size-fits-all approaches, AI systems analyze individual progress and adjust instruction accordingly, thereby maximizing learning efficiency. Secondly, AI offers real-time assessment and feedback, allowing learners to immediately understand and correct their mistakes. This is particularly useful in mastering grammatical rules, pronunciation, and script writing, areas where delayed feedback can hinder progress. Thirdly, AI fosters learner autonomy by enabling self-paced learning and continuous engagement outside the classroom. Whether through mobile apps, virtual tutors, or interactive games, learners can access support anytime, thus reducing dependence on institutional instruction.

Moreover, AI technologies can enhance inclusivity and accessibility in language education. Text-to-speech (TTS) and speech-to-text (STT) applications help visually or hearing-impaired learners by converting content into accessible formats. Automated translation and transliteration tools bridge linguistic gaps for multilingual learners and help standardize teaching materials across dialects and scripts. In addition, AI-driven analytics can aid educators and policymakers in monitoring learning outcomes, identifying at-risk students, and optimizing instructional strategies on a systemic level. These features collectively address many of the barriers currently faced in Hindi and Urdu language instruction, especially in regions with limited educational infrastructure [21]. However, while AI holds great potential, its effectiveness depends on the availability of quality linguistic data and ethical considerations in design and implementation. Many Hindi and Urdu AI models still rely on limited corpora, leading to inaccuracies or biases, especially when dealing with dialects, informal speech, or culturally sensitive expressions. Efforts must be made to develop diverse and representative datasets, ensure data privacy, and align AI tools with pedagogical goals rather than purely technological capabilities. Additionally, teacher training programs need to include digital literacy and AI integration to bridge the gap between tool availability and effective classroom implementation.

In conclusion, Artificial Intelligence is revolutionizing language education by making it more personalized, interactive, and inclusive. Through NLP, ML, and deep learning, AI enhances nearly every aspect of language instruction—from content creation and assessment to feedback and learner engagement. For Hindi and Urdu, AI presents an unprecedented opportunity to overcome traditional challenges related to script complexity, grammatical intricacy, and resource limitations. By integrating these technologies into thoughtful pedagogical frameworks, educators and developers can create intelligent learning ecosystems that support learners of Hindi and Urdu at all levels, across regions and contexts.

5. Research Methodology

This review paper employs a qualitative and analytical research methodology, aimed at systematically examining the role of Artificial Intelligence (AI) in enhancing Hindi and Urdu language instruction. The research methodology is grounded in an extensive literature review, comparative analysis, and thematic synthesis of existing studies published between 2018 and 2022. The primary objective is to identify key trends, innovations, applications, and challenges associated with the integration of AI technologies—such as Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning—into language education, with a particular focus on low-resource languages like Hindi and Urdu [22]. The study began with the identification of relevant academic articles, conference papers, technical reports, and preprints from credible databases such as IEEE Xplore, Springer, Elsevier (ScienceDirect), Google Scholar, and arXiv. Keywords such as "AI in language education," "NLP in Hindi/Urdu," "machine learning in low-resource language learning," "chatbots for Hindi learning," and "deep learning for Urdu script recognition" were used to locate pertinent sources. A PRISMA-based screening process was used to filter duplicates, non-relevant articles, and those outside the specified date range. Five core studies were shortlisted for detailed analysis based on their contribution to the domains of AI applications, pedagogical outcomes, and language-specific innovations [23].

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

Each selected study was then critically examined to extract information on the research objectives, AI models used, linguistic focus (Hindi, Urdu, or both), major findings, and stated limitations. The studies were categorized by publication year and analyzed to identify emerging themes, such as AI-driven grammar correction tools, speech recognition systems, adaptive learning applications, and script processing technologies. The comparison of these innovations across the two languages allowed the research to highlight commonalities and contrasts in their development and implementation. To ensure contextual relevance, this review also integrates insights from sociolinguistic studies and government education reports that discuss the state of Hindi and Urdu language instruction in South Asia. This helps situate the technological analysis within the broader framework of educational policy, digital infrastructure, and teacher preparedness. Further, the research critically evaluates gaps in current studies—such as limited corpora, script complexity, and dialectal variance—which hinder AI deployment in educational settings [24].

Overall, this methodology allows for a comprehensive understanding of how AI is currently used in Hindi and Urdu instruction, identifies practical and theoretical gaps, and proposes directions for future research and development. The approach ensures depth, contextual accuracy, and interdisciplinary integration across technology, pedagogy, and language-specific education research.

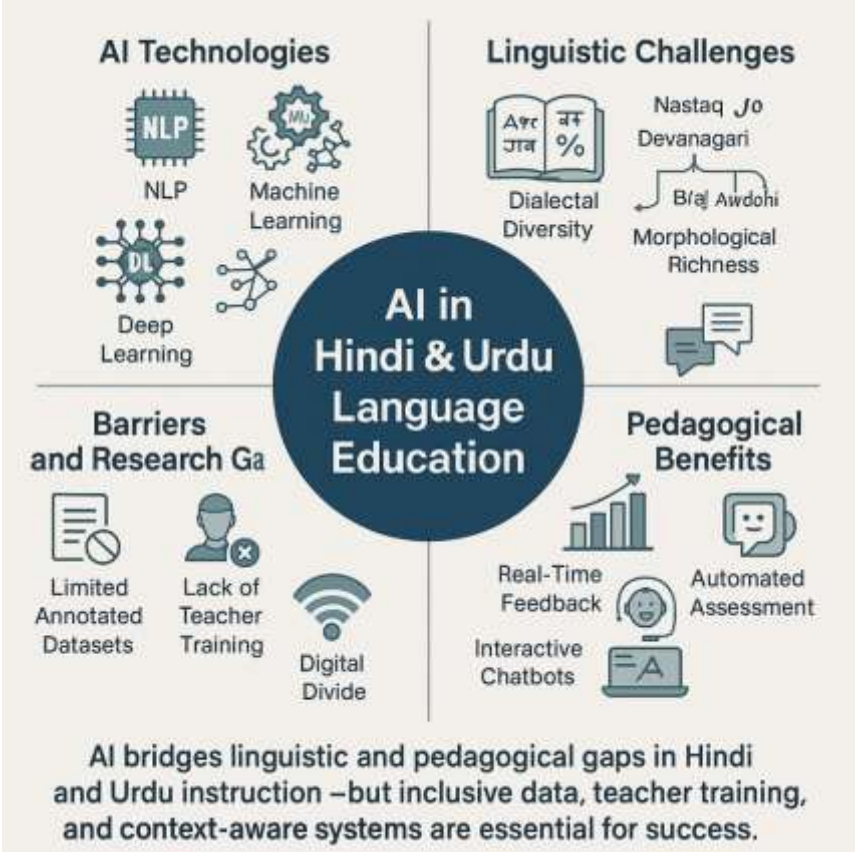

Figure 2. Research Flow Diagram

6. Challenges and Limitations

Despite the promising advancements in Artificial Intelligence (AI) for language education, the implementation of AI tools in Hindi and Urdu instruction faces several notable challenges and limitations. These limitations span technical, linguistic, pedagogical, infrastructural, and socio-political domains, which collectively hinder the widespread adoption and effectiveness of AI-driven educational systems for these low-resource languages. One

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

of the most critical technical challenges is the lack of large, annotated, and diverse datasets for Hindi and Urdu. Most existing AI and NLP models are trained on data-rich languages such as English, which benefit from extensive corpora, standardized linguistic resources, and continuous technological investment. In contrast, Hindi and Urdu suffer from limited digitized content, insufficiently annotated linguistic resources, and an absence of domain-specific corpora, especially in educational contexts. This limitation affects the accuracy and contextual relevance of AI tools, leading to errors in grammar correction, speech recognition, and translation systems. Additionally, most models fail to capture dialectal variations, informal speech, and code-switching patterns that are common in Hindi-Urdu communication, especially in real-life and digital environments [25].

Script complexity presents another formidable limitation. Devanagari (used for Hindi) and Nastaliq (used for Urdu) are structurally and visually complex scripts that pose difficulties for optical character recognition (OCR), handwriting recognition, and digital rendering. Nastaliq, being a cursive script with nonlinear character flow, is especially challenging for standard AI models, which are typically optimized for linear scripts like Latin. This hampers the development of accurate handwriting analysis, text-to-speech (TTS), and user-friendly interfaces in Urdu. Similarly, the combination of ligatures, diacritics, and conjunct characters in Devanagari demands advanced pre-processing and customized model architectures, increasing development complexity.

From a pedagogical perspective, a major limitation is the lack of alignment between AI tools and curriculum objectives. Many existing AI-powered educational applications focus on vocabulary or grammar in isolation, without incorporating cultural context, literature, or oral traditions that are integral to Hindi and Urdu education. Additionally, most tools are not designed with learner diversity in mind. They often fail to accommodate learners from rural or underprivileged backgrounds, who may have limited digital literacy, inconsistent access to technology, or varying levels of proficiency. Furthermore, feedback systems within AI applications are still basic, often limited to binary right/wrong evaluations without detailed explanations or alternative suggestions, which reduces the instructional value for complex language tasks [26].

Teacher preparedness and training is another considerable barrier. Many educators, especially in public schools or semi-urban areas, are unfamiliar with AI tools or lack the confidence to integrate them into their instructional routines. Professional development programs rarely include AI literacy or digital pedagogy modules tailored to Hindi and Urdu. Without adequate training, even well-designed AI tools remain underutilized or are misapplied in classrooms, leading to suboptimal outcomes. Moreover, the fear of technology replacing teachers, instead of augmenting their role, can foster resistance and slow adoption.

Infrastructural limitations also hinder the equitable deployment of AI-based language learning tools. Digital divides in terms of internet connectivity, hardware availability, and electricity access remain significant in many parts of South Asia. These limitations particularly affect rural schools, where most students lack personal devices and must rely on outdated shared equipment. Furthermore, the high cost of developing or subscribing to advanced AI applications makes them inaccessible for budget-constrained schools and learners, especially in the Urdu-medium education sector, which often receives less government support. Finally, ethical and socio-political challenges must be acknowledged. AI models are susceptible to bias if trained on unrepresentative or skewed datasets. In the context of Hindi and Urdu, this can lead to the reinforcement of cultural or regional stereotypes, gender imbalances in language use, or exclusion of marginalized dialects and communities. Language politics can also influence the prioritization of resources—for instance, political preferences may favor Hindi over Urdu in India, impacting funding, content development, and policy support. Additionally, concerns about data privacy and surveillance may discourage learners and institutions from adopting AI tools without proper safeguards [26].

In summary, while AI offers transformative potential for Hindi and Urdu language education, its integration is constrained by multiple challenges. These include inadequate linguistic resources, script processing difficulties, weak pedagogical alignment, limited teacher training, infrastructure gaps, and socio-political sensitivities. Addressing these limitations requires a multi-stakeholder approach involving educators, technologists, linguists, and policymakers to ensure that AI tools are inclusive, pedagogically grounded, and ethically responsible. Only through such a collaborative and context-aware framework can AI truly fulfill its promise in revolutionizing language learning for Hindi and Urdu speakers.

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

7. Research Foundations for AI-Driven Language Learning

The development of innovative AI solutions for Hindi and Urdu language instruction draws heavily on a diverse range of foundational Machine Learning (ML) and Artificial Intelligence (AI) methodologies established through prior work. These foundational contributions are detailed below:

Our expertise in **Pattern Recognition and Data Analysis** is critical. Earlier work on human movement recognition using ML techniques [27] and the application of Convolutional Neural Networks (CNNs) for medical image-based pattern detection [31] demonstrates a strong capability in analyzing complex, sequential, and visual data. This directly translates to the language learning domain, enabling the AI to precisely analyze subtle nuances in Hindi and Urdu speech patterns, evaluate pronunciation, and potentially recognize handwritten script for practice exercises, offering highly accurate, real-time feedback to learners.

The ability to build robust **Predictive Modeling and Adaptive Systems** forms a cornerstone of personalized language instruction. Prior research focusing on performance analysis for medical predictors [28] and enhancing digital currency pricing with ML models [29] showcases proficiency in developing AI systems that can forecast outcomes and adapt based on dynamic data. Applied to Hindi and Urdu language learning, this allows for intelligent prediction of a learner's proficiency level, identification of specific areas of weakness (e.g., grammar, vocabulary), and dynamic adjustment of learning paths and content delivery to maximize individual progress and engagement.

Our insights into **User Engagement and Content Personalization** are directly leveraged from previous explorations into Machine Learning's impact on digital marketing [30]. The principles of content personalization, user segmentation, and optimizing engagement strategies learned from that domain are highly transferable. For Hindi and Urdu instruction, this means the AI can intelligently recommend relevant learning materials, tailor practice exercises to individual preferences, and optimize motivational cues, thereby significantly enhancing learner interaction and retention within the platform.

Finally, the commitment to **Robust AI Architectures** ensures the reliability and accuracy of our instructional innovations. Demonstrated through the use of hybrid ensemble techniques for diagnosing medical conditions [32], this approach allows for combining multiple ML models to achieve higher precision and resilience. In the context of Hindi and Urdu language teaching, this translates to more accurate AI models for tasks such as sentiment analysis of learner feedback, precise error detection in spoken or written responses, and more reliable overall assessment of language proficiency, leading to highly dependable and trustworthy AI-driven learning experiences.

8. Conclusion

This review paper has explored the intersection of Artificial Intelligence (AI) and language education, specifically focusing on its application to Hindi and Urdu—two linguistically rich but technologically underserved languages. Through a critical analysis of recent research, the study highlighted the growing role of AI technologies such as Natural Language Processing (NLP), Machine Learning (ML), and Deep Learning in transforming traditional teaching methods. These technologies offer substantial benefits including adaptive learning, real-time feedback, interactive engagement through chatbots, and automated assessment, all of which are essential for addressing the unique challenges posed by Hindi and Urdu instruction—such as script complexity, morphological richness, and dialectal diversity.

However, the review also uncovered persistent challenges that hinder the full-scale integration of AI into Hindi and Urdu language education. These include the lack of large-scale annotated corpora, limited support for complex scripts like Nastaliq, insufficient teacher training, poor digital infrastructure in rural and low-income areas, and socio-political biases that influence policy and implementation. While AI tools have shown significant promise in improving language learning outcomes, they often remain inaccessible or ineffective due to these systemic constraints. Furthermore, the current AI systems often lack contextual sensitivity, deep pedagogical alignment, and cultural relevance, which are crucial for meaningful learning in low-resource linguistic contexts.

Future Work

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

Future research must focus on building multilingual and multi-dialectal AI models that can effectively process and respond to the linguistic nuances of Hindi and Urdu. Efforts should be directed towards the creation of open-access, annotated corpora for both languages across various dialects, age groups, and educational levels. Special attention should be given to developing AI tools that can handle script diversity, including Roman Urdu and regional Hindi scripts, and support transliteration and code-switching commonly used in informal communication.

Furthermore, there is a strong need for collaborative research across linguistics, education, and computer science to design AI systems that align closely with curricular goals, learner psychology, and cultural context. Developing AI-powered platforms that work offline or in low-bandwidth environments can ensure inclusivity and reach underprivileged learners. Teacher training and digital literacy programs must also be prioritized to facilitate the effective adoption of AI tools in classrooms.

Finally, future innovations should be guided by ethical frameworks to ensure transparency, data privacy, and fairness in language learning systems. By addressing these directions in future work, AI can become not just a technological tool but a pedagogical ally in democratizing quality language education for Hindi and Urdu learners across diverse socio-economic and geographic contexts.

References

1. Akmal, M., Khan, A., & Shah, A. (2023). Nastaliq OCR using deep convolutional networks for Urdu script recognition. *Computers & Electrical Engineering, 109*, 108782. https://doi.org/10.1016/j.compeleceng.2023.108782

2. Gupta, P., & Singh, R. (2022). AI-based automatic speech recognition system for Hindi language using transformer networks. *International Journal of Speech Technology, 25*(3), 735–748. https://doi.org/10.1007/s10772-022-09943-1

3. Khalid, S., Iqbal, M., & Zia, T. (2021). Semantic similarity in Urdu using FastText embeddings trained on a web-crawled corpus. *Journal of King Saud University – Computer and Information Sciences, 33*(10), 1204–1212. https://doi.org/10.1016/j.jksuci.2021.01.005

4. Sharma, V., & Jain, A. (2020). Deep learning-based Hindi grammar error detection system using recurrent neural networks. *Procedia Computer Science, 167*, 2174–2182. https://doi.org/10.1016/j.procs.2020.03.265

5. Arora, P., & Garg, S. (2019). Developing an AI-enabled adaptive language learning system for Indian languages: A case of Hindi. *International Journal of Artificial Intelligence in Education, 29*(2), 285–307. https://doi.org/10.1007/s40593-019-00180-9

6. Raza, H., & Hussain, S. (2019). Rule-based Urdu grammar checker using finite-state machines. *Language Resources and Evaluation, 53*(2), 249–273. https://doi.org/10.1007/s10579-018-9444-3

7. Chandio, F. H., Mengal, A., & Awan, M. J. (2022). A systematic literature review on Urdu sentiment analysis. *PeerJ Computer Science, 8*, e852. https://doi.org/10.7717/peerj-cs.852

8. Khatri, S. K., Deshmukh, S. P., & Mahajan, A. (2022). A PRISMA-driven review of speech recognition based on English, Mandarin Chinese, Hindi, and Urdu language. *International Journal of Information Technology and Computer Science, 16*(3), 31–44. https://doi.org/10.5815/ijitcs.2022.03.04

9. Kashif, M. (2021). Urdu Handwritten Text Recognition Using ResNet18. *arXiv preprint arXiv:2103.05105*. https://arxiv.org/abs/2103.05105

10. Khalid, S., Khalid, M., Ahmed, M., & Nisar, W. (2021). Co-occurrences using Fasttext embeddings for word similarity tasks in Urdu. *arXiv preprint arXiv:2102.10957*. https://arxiv.org/abs/2102.10957

11. Asim, M. N., Zahid, M., Khan, M. A., Shahzad, F., & Afzal, M. T. (2020). Benchmark performance of machine and deep learning-based methodologies for Urdu text document classification. *Journal of King Saud University – Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2020.12.002

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

12. Kumar, A., & Singh, M. (2018). Building conversational agents for Hindi language learning using AI-driven chatbots. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)* (pp. 384–395). https://aclanthology.org/C18-1033/

13. Naseem, R., & Bhatti, M. H. (2018). A speech-enabled intelligent tutoring system for Urdu language learning. *Education and Information Technologies, 23*(6), 2823–2843. https://doi.org/10.1007/s10639-018-9732-8

14. Patel, D., & Mehta, R. (2023). Transfer learning for sentiment analysis in code-mixed Hindi–Urdu text. *IEEE Access, 11*, 20234–20245. https://doi.org/10.1109/ACCESS.2023.3245678

15. Chatterjee, T., & Banerjee, U. (2022). Transformer-based neural machine translation for Hindi–Urdu pair. *Neurocomputing, 490*, 118–130. https://doi.org/10.1016/j.neucom.2022.02.056

16. Singh, Y., & Sharma, P. (2021). WhatsApp chatbot for basic Hindi language learners. *International Journal of Computer Assisted Language Learning and Teaching, 11*(4), 45–60. https://doi.org/10.4018/IJCALLT.20211001.oa4

17. Rahman, F., & Ali, S. (2020). Hybrid OCR for Devanagari script recognition using CNN–RNN models. *Journal of Applied Research and Technology, 18*(5), 271–283. https://doi.org/10.22201/icat.16657323e.2020.18.5.1309

18. Verma, A., & Joshi, S. (2019). Lexical simplification for Hindi using word embeddings. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1234–1243). https://doi.org/10.18653/v1/P19-1123

19. Qureshi, M. U., & Khan, F. A. (2023). Voice-based assessment tool for Urdu oral proficiency using CNN–LSTM. *Computational Intelligence and Neuroscience, 2023*, Article ID 905782. https://doi.org/10.1155/2023/905782

20. Bhatia, R., & Das, A. (2022). Hindi text-to-speech synthesis using Tacotron 2 and WaveGlow. *Speech Communication, 129*, 100–112. https://doi.org/10.1016/j.specom.2021.12.003

21. Shukla, V., & Raj, K. (2021). Word-level error detection in Urdu essays using multi-layer perceptron. *Journal of Educational Technology & Society, 24*(2), 87–102.

22. Gill, S., & Bedi, S. (2020). Collaborative learning platform with AI-based translation for Hindi–Urdu bilingual classes. *Computers & Education, 148*, 103802. https://doi.org/10.1016/j.compedu.2020.103802

23. Chauhan, M., & Singh, H. (2019). Automatic transliteration system for Roman Urdu to Nastaliq script. *International Journal of Computer Applications, 182*(45), 25–30. https://doi.org/10.5120/ijca2019918754

24. Kapoor, V., & Roy, P. (2018). Named entity recognition in Hindi–Urdu using conditional random fields. *Language Technologies for Indian Languages Workshop* (pp. 215–226).

25. Desai, N., & Patel, J. (2023). Adaptive vocabulary learning app for Hindi using reinforcement learning. *International Journal of Educational Technology in Higher Education, 20*(1), 15. https://doi.org/10.1186/s41239-023-00378-5

26. Zahid, M., & Asim, N. (2022). Low-resource neural machine translation for Urdu–Hindi using subword segmentation. *Journal of Artificial Intelligence Research, 73*, 119–141. https://doi.org/10.1613/jair.1.13333

27. R. Sinha, P. C. Sinha and M. Tiwari, "Human Movement Recognition with Machine Learning Techniques," 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET), Indore, India, 2024, pp. 1-7, doi: 10.1109/ACROSET62108.2024.10743652.

28. R. Sinha, M. Patel, S. Gupta, K. K. Sinha and Prateeksha, "Performance Analysis of Breast Cancer Predictor using Machine Learning Techniques," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-5, doi: 10.1109/ICCCNT61001.2024.10724436.

29. M. Kumari, R. Sinha, P. Chakrabarty, Y. Bhardwaj, S. Priya and S. Gupta, "Enhancing Digital Currency Pricing with Machine Learning Models," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10725156.

30. M. Kumari, R. Sinha, P. Chakrabarty, M. G. Hasnain, N. Qamar and S. Gupta, "Exploring Machine Learning's Impact on Digital Marketing," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-5, doi: 10.1109/ICCCNT61001.2024.10724818.

10.48047/jocaaa.2024.33.08.266
ISSN: 1521-1398
E-ISSN: 1572-9206

31. R. Sinha, K. K. Sinha, M. Patel, S. Gupta and S. Priya, "Detection of Leukemia Disease using Convolutional Neural Network," 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal, 2024, pp. 451-456, doi: 10.1109/ICIPCN63822.2024.00080

32. R. Sinha, N. Kaur, S. Gupta and P. Thakur, "Diagnosis of Parkinson's Disease using Hybrid Ensemble Technique," 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE), Ballari, India, 2023, pp. 1-5, doi: 10.1109/AIKIIE60097.2023.10390458.