

Enhancing Machine Translation and Speech Recognition for Low-Resource Indian Languages

Barkha Sahu

Assistant Professor, Department of Computer Science and Engineering
Institute of Engineering & Science, IPS Academy, Indore, India
bsahuscs@gmail.com

Abstract— Machine Translation (MT) and Automatic Speech Recognition (ASR) systems have made significant strides globally; however, low-resource Indian languages remain underrepresented in these technological advancements. The complex linguistic landscape of India, characterized by diverse dialects, code-mixing, and limited annotated data, presents unique challenges for building efficient MT and ASR systems. Recent developments in neural machine translation, transformer models, and time-delay neural networks have shown promise in improving translation accuracy and speech recognition, particularly for Hindi and other regional languages. Despite these efforts, there is a need for more comprehensive datasets, especially in specialized domains like legal and medical terminology, to improve translation fidelity. Moreover, pooling strategies and adaptive modeling in ASR systems must be refined to handle noisy environments and dialect variations effectively. Future work should focus on creating rich, multilingual corpora, advancing transfer learning techniques, and fostering interdisciplinary collaboration. Such approaches will help bridge the digital language divide, enabling more inclusive language technologies for India's linguistically diverse population.

Keywords : Low-Resource Languages , Machine Translation , Speech Recognition , Indian Languages , Neural Networks, Code-Mixing.

I. INTRODUCTION

India is a nation renowned for its linguistic diversity, with 22 constitutionally recognized official languages and over 1,600 dialects spoken across its vast geography. Despite this richness, the advancement of language technologies, particularly **Machine Translation (MT)** and **Automatic Speech Recognition (ASR)** systems, has been uneven, predominantly benefiting high-resource languages like English, Mandarin, and Spanish [1]. Indian languages, especially those categorized as low-resource, remain on the periphery of technological progress due to insufficient linguistic data, complex syntax, regional dialects, and widespread code-mixing practices. These limitations have created a pressing need to enhance both MT and ASR systems tailored specifically for the Indian linguistic ecosystem to ensure equitable access to digital services, education, and information dissemination [2].

Machine Translation has witnessed a paradigm shift from **Phrase-Based Statistical Machine Translation (PB-SMT)** to **Neural Machine Translation (NMT)**, leading to substantial improvements in translation fluency and contextual accuracy. However, these benefits have not been fully realized for Indian languages due to the scarcity of large, domain-specific parallel corpora and annotated datasets [3]. The diversity in grammar, semantics, and script across languages like Hindi, Marathi, Tamil, and Telugu adds to the complexity of developing universally efficient MT systems. Studies have highlighted significant terminology translation errors in both PB-SMT and NMT, particularly when dealing with specialized vocabulary in fields such as law, medicine, and science. The absence of standardized datasets for terminology [4] translation further exacerbates this challenge, leaving a critical research gap that

must be addressed through the creation of expansive, quality datasets and domain-specific lexicons.

Similarly, ASR systems in India face multifaceted challenges stemming from linguistic plurality and socio-cultural factors. Conventional ASR systems have traditionally relied on **Gaussian Mixture Models-Hidden Markov Models (GMM-HMM)**, but recent advancements have shifted focus to **deep learning-based models**, notably **Time-Delay Neural Networks (TDNN)** and **Convolutional Neural Networks (CNN)**. These models have shown performance improvements in speech recognition [5] tasks for Hindi and other regional languages. However, they often struggle with the variability introduced by dialects, accents, and environmental noise common in real-world scenarios. Furthermore, pooling techniques, essential for feature extraction in CNN architectures, need further exploration to optimize performance under diverse conditions, including clean and noisy speech data.

Another significant obstacle in enhancing ASR and MT for low-resource Indian languages is the prevalence of **code-mixing**, where speakers interchangeably use words and phrases from different [6] languages within the same conversation. This phenomenon, widespread in Indian society, especially between Hindi and English, poses considerable difficulties for language models trained predominantly on monolingual datasets. Addressing this requires innovative modeling approaches that can effectively understand and process multilingual and code-mixed data [7].

The limited availability of annotated corpora, standardized evaluation metrics, and computational resources tailored for Indian languages has further hindered progress. While some initiatives like the **Science Answer Assessment (ScAA)** dataset for Hindi and Marathi have emerged, they are insufficient to cover the vast linguistic landscape

comprehensively. Consequently, there is an urgent need to develop more inclusive datasets [8] encompassing diverse dialects, sociolects, and domain-specific content to train and benchmark MT and ASR models effectively.

To bridge these gaps, future research should prioritize **transfer learning, semi-supervised learning, and multilingual neural modeling**, which allow models trained on high-resource languages to be adapted for low-resource ones with limited data. Interdisciplinary collaboration among linguists, computer scientists, and policymakers will be critical in creating sustainable frameworks [9] for data collection, annotation, and model development. Additionally, government and private sector partnerships can drive initiatives to build **open-source datasets** and tools that democratize access to language technologies [10].

In conclusion, enhancing machine translation and speech recognition for low-resource Indian languages is imperative for fostering digital inclusivity and preserving linguistic diversity. By leveraging advanced neural architectures, creating robust datasets, and adopting culturally contextualized research methodologies, India can develop scalable solutions that cater to its multilingual population, ensuring that language is not a barrier to accessing the benefits of the digital age.

II. LITERATURE REVIEW

This chapter explores diverse subdisciplines within linguistics and applied linguistics, including descriptive linguistics, sociolinguistics, sociology of language, bilingualism, second language acquisition, foreign language teaching, and stylistics. It emphasizes how multilingualism has been integral to the Indian civilizational narrative through contact, conflict, rivalry, and assimilation. The chapter specifically examines Indian English, especially 'Educated Indian English,' by synthesizing published descriptions, literary analysis, and empirical data from spoken and written sources, including corpora. This comprehensive perspective reveals the hybrid nature of Indian English and its sociolinguistic significance within India's multilingual context (Sridhar, 2020) [1].

The study by Zaki-Hindi et al. (2020) investigates Ultra-Reliable Low-Latency Communication (URLLC) in Industry 4.0, focusing on uplink transmissions across licensed and unlicensed spectrums. The authors propose a cost-efficient transmission policy prioritizing unlicensed spectrum before shifting to licensed bands within the delay constraint. They initially examine single-tenant scenarios to minimize licensed bandwidth costs and extend their analysis to multi-tenant scenarios using game theory, addressing competition in shared unlicensed spectrum. They prove the existence of pure Nash equilibria and quantify the price of anarchy, offering insights into optimizing URLLC transmissions in industrial multi-tenant settings (Zaki-Hindi et al., 2020)[2].

Haque et al. (2020) conduct a comparative study on terminology translation between phrase-based statistical machine translation (PB-SMT) and neural machine translation (NMT) in English-Hindi and Hindi-English directions. Given the absence of a gold standard for domain-specific

terminology translation, they develop a legal corpus-based evaluation test set. The study introduces an error typology specific to terminology translation, classifying errors manually and benchmarking both MT approaches. Their findings expose the respective strengths and weaknesses of PB-SMT and NMT in handling specialized terms, providing a foundation for future advancements in machine translation quality assessment (Haque et al., 2020)[3].

Mubarki (2020) explores representations of the male body in Hindi cinema, paralleling its female counterpart as a site of nationalistic and hegemonic expression. The ideal male physique has evolved from Gandhi-inspired lean frames of the colonial era, to well-fed socialist bodies of the Nehruvian period, culminating in the muscular, hyper-masculine physiques aligned with Hindutva ideologies. This cinematic portrayal reflects the shifting aspirations and political narratives of India across decades, linking body aesthetics with national discourse and identity construction in popular culture (Mubarki, 2020)[4].

Iyer (2020) analyzes Tamil and Hindi films from the 1950s starring Vyjayanthimala, focusing on the cultural transmission of Bharatanatyam dance into Bombay cinema. Directed by M.V. Raman and produced by AVM Productions, these films introduced classical dance gestures, influencing costumes, narratives, and aesthetics. This cultural cross-pollination from Tamil to Hindi cinema reshaped narrative styles and contributed to evolving industry hierarchies. The star's embodiment of Bharatanatyam highlighted regional influences in shaping the national cinematic identity, enriching Bombay cinema with South Indian cultural elements (Iyer, 2020)[5].

Strauss (2020) examines the global journey of yoga, tracing its transformation from an Indian spiritual practice to an international phenomenon. Originating from Swami Vivekananda's presentation at the 1893 Chicago Parliament of World Religions, yoga adapted to Western sensibilities, evolving into various modern practices. The book discusses how global lifestyle values, commercialization, and cross-cultural exchanges have redefined yoga while maintaining its philosophical roots. Yoga's popularity reflects cultural flows, producing new meanings across boundaries while retaining its connection to Indian spiritual traditions (Strauss, 2020)[6].

Boruah (2020) documents the language loss among Gond tribes, India's second-largest tribal group, and community efforts for revitalization. Based on field data from Madhya Pradesh and Chhattisgarh, the study reveals Gondi's vulnerability as classified by UNESCO, with restricted use despite being widely spoken. Linguistic analysis shows Gondi's convergence with dominant regional languages like Hindi and Chhattisgarhi. The community's revitalization efforts face challenges such as socio-economic changes and policy neglect. The paper underscores the need for targeted language preservation initiatives to sustain linguistic diversity (Boruah, 2020)[7].

Bhatnagar et al. (2020) validate the Hindi version of the Difficulties in Emotion Regulation Scale (DERS-H), addressing debates on its factor structure. Studying 434 participants, the authors identify psychometric weaknesses in the Awareness dimension, ultimately proposing a modified five-factor model with better fit, reliability, and validity. Their results show that the refined DERS-H provides a robust tool for assessing emotion regulation in the Indian cultural context, maintaining structural consistency across gender and age groups (Bhatnagar et al., 2020)[8].

Siddiqui and Siddiqui (2020) explore the socio-cultural elevation of Urdu among North and Deccan Indian Muslims. Urdu, historically linked with the elite and the Muslim League, serves as a symbol of identity and status within the community. The study notes the nostalgic revival of Persianized Nastaliq Urdu as a hegemonic language form, sustained through the Hindi-Urdu controversy under British rule. This linguistic preference reinforces elitism and distinctiveness among Muslims, cementing Urdu's role in cultural and political identity (Siddiqui & Siddiqui, 2020)[9].

De Silva et al. (2020) challenge classical language dynamics models by presenting a case where Hindi, a lower-status language, outcompetes English in India. Analysis of Indian census data (1961–2011) reveals a saturation in Hindi-English bilingualism, while monolingual Hindi speakers rise sharply. Their novel model captures this unique linguistic shift, contrasting with global trends where English typically dominates. This suggests socio-political and cultural factors in India bolster Hindi's prominence over English, offering new insights into language competition dynamics (De Silva et al., 2020)[10].

Passricha and Aggarwal (2020) compare pooling strategies within convolutional neural networks (CNN) for Hindi automatic speech recognition (ASR). Their study evaluates various pooling techniques—max, average, stochastic—on Hindi speech datasets under clean and noisy conditions. Results indicate max pooling excels in clean speech, while stochastic pooling suits noisy environments. This work enhances understanding of CNN mechanics in ASR, addressing the "black-box" nature of these models, and offers guidance on optimizing pooling strategies for speech recognition tasks (Passricha & Aggarwal, 2020)[11].

Sohail (2020) reviews the role of Sufis and saints in the origin and evolution of Urdu in the Indian subcontinent. While Muslims arrived as traders, conquerors, and preachers, Sufis significantly influenced Urdu's development through their spiritual and social outreach. Their contributions enriched Urdu's lexicon and literary heritage, fostering a syncretic cultural identity. This historical review highlights the spiritual roots of Urdu and its evolution into a distinctive language that reflects the region's composite culture (Sohail, 2020)[12].

Gupta (2020) examines the Hindi print-public sphere's portrayal of indentured women between 1917 and 1920, using periodicals like *Stri Darpan* and literary works by poets like

Maithili Saran Gupta. These narratives, though claiming to highlight women's plight, often sensationalized their experiences. The discourse facilitated middle-class women's activism for overseas sisters while avoiding domestic gender reforms. This study uncovers the complexities of gender, print culture, and coloniality in shaping public opinion about women's experiences in indenture (Gupta, 2020)[13].

Osmani et al. (2020) propose a structured program to empower female entrepreneurs through co-innovation platforms and ICT tools. The model includes stakeholder engagement, entrepreneurship training, capacity building, pilot-testing, and international collaboration. The initiative seeks to enhance entrepreneurial activities by fostering innovation, mentorship, and venture capital access. This framework aims to bridge gender gaps in entrepreneurship, facilitating knowledge exchange and supporting women-led start-ups globally (Osmani et al., 2020)[14].

Kumar and Mehta (2020) validate the Receptive Expressive Emergent Language Test-3 (REELT-3) for Hindi-speaking children with cochlear implants. Their study confirms the test's reliability and validity, demonstrating strong correlations between language scores and chronological age. The adaptation ensures that REELT-3 accurately assesses language outcomes in Hindi-speaking post-implantation children, supporting effective clinical evaluation and therapeutic planning for language development in this population (Kumar & Mehta, 2020)[15].

Li et al. (2020) investigate the thermal stability of methyltrimethoxysilane (MTMS) silica aerogels (MSA), noting superior thermal resistance with onset and peak temperatures of 417°C and 476°C. Through pyrolysis studies, they identify decomposition stages and associated activation energies. The findings contribute to developing safer, high-performance nanoporous materials for thermal insulation, offering guidance on reducing thermal hazards in hydrophobic silica aerogels (Li et al., 2020)[16].

Rahman et al. (2020) evaluate shrinkage-mitigating concrete mixtures in scaled bridge decks using technologies like shrinkage-compensating cement, internal curing, and shrinkage-reducing admixtures. The study compares these with standard concrete, revealing significant reductions in shrinkage and improved performance. Their results inform infrastructure engineering, presenting viable solutions for enhancing bridge deck durability and longevity against cracking (Rahman et al., 2020)[17].

Agarwal et al. (2020) introduce the Science Answer Assessment (ScAA) dataset for automatic short answer grading (ASAG) in Hindi and Marathi, addressing challenges in low-resource languages. The dataset includes over 10,000 labeled responses from children aged 8-14. Benchmarking various ASAG methods, the study establishes a foundation for advancing NLP in Indian regional languages and enhancing educational assessments through AI (Agarwal et al., 2020)[18].

Virk et al. (2020) discuss Ayurveda's conceptualization of yoga as the union of body, mind, and soul, promoting psychosomatic balance for health. This traditional understanding aligns with modern findings that link yoga to natural dopamine production, stress reduction, and holistic wellness. Yoga's ancient framework, encompassing Yama, Niyama, Asana, and other limbs, offers a comprehensive system for maintaining mental and physical equilibrium (Virk et al., 2020)[19].

Khatri (2020) examines linguistic codes in Hindi cinema, noting how regional dialects and Creole forms create a "Semantic Enclave." Analyzing films rooted in regional milieus, the study reveals how cinematic language mediates rural-urban divides and critiques mainstream development narratives. The absence of standardized subtitles further alienates regional linguistic identities within Hindi cinema (Khatri, 2020)[20].

Kumar and Aggarwal (2020) enhance Hindi automatic speech recognition (ASR) systems using Time Delay Neural Networks (TDNN), achieving an 11% performance boost over traditional GMM-HMM models. Further gains were realized through i-vector adaptation and language model interpolation, addressing the resource limitations in Hindi ASR development and improving speech recognition capabilities (Kumar & Aggarwal, 2020)[21].

Mubarki (2020) critiques Hindi cinema's treatment of culinary diversity, particularly the marginalization of meat consumption. Since the 1970s, Bollywood has visually excluded meat-eating, reinforcing a collectivized cultural identity tied to vegetarianism and nationalistic purity. This erasure reflects broader socio-cultural anxieties about "Otherness" in India's cinematic representations (Mubarki, 2020)[22].

III. RESEARCH GAPS

1. **Multilingualism and Linguistic Diversity in India:** Although multilingualism is deeply embedded in India's socio-cultural fabric, existing research lacks comprehensive empirical studies on the evolving dynamics between regional languages and English, particularly in urban educational contexts. Most studies focus on descriptive accounts without integrating longitudinal data to assess language shift, maintenance, or hybridization processes in real educational settings.
2. **Language Preservation in Tribal Communities:** Current studies on language loss, such as in the Gondi community, highlight vulnerability but provide limited scalable frameworks for revitalization across multiple states. There is a research gap in formulating sustainable language preservation models that leverage technology, policy interventions, and educational reforms to revive endangered languages in India.

3. **Emotion Regulation Measurement Tools in Indian Contexts:** While the DERS-H scale has been validated, there is insufficient cross-cultural psychometric research comparing its application across various socio-economic and linguistic groups within India. The scalability of such tools for non-clinical populations or in diverse cultural contexts remains underexplored.
4. **Machine Translation and Terminology in Low-Resource Languages:** Machine Translation (MT) for Indian languages, especially for domain-specific terminology, remains underdeveloped. There is a clear gap in building comprehensive, annotated datasets across different Indian languages for MT performance, particularly in specialized domains like legal, medical, and technical texts.
5. **Hindi Automatic Speech Recognition (ASR) and Pooling Techniques:** Despite improvements using CNN and TDNN models for Hindi ASR, research has yet to fully explore advanced neural architectures like transformers for speech tasks in noisy and multi-dialectal Indian environments. Comparative analysis of such models against existing pooling methods is largely missing.
6. **Cultural Representation in Hindi Cinema:** The studies reveal selective linguistic and cultural representations in Hindi cinema, yet there is limited research on how these representations impact language perception, usage, and regional identities among audiences. Additionally, the lack of standardized subtitles and linguistic inclusivity in cinema remains an overlooked area.
7. **Gender, Entrepreneurship, and Technology:** Existing research proposes frameworks for fostering women entrepreneurship through co-innovation platforms, yet empirical validation across different regions in India remains scarce. There is a research gap in understanding region-specific barriers and ICT adoption levels among aspiring female entrepreneurs in semi-urban and rural areas.
8. **Automatic Short Answer Grading in Low-Resource Languages:** Although datasets like ScAA for Hindi and Marathi exist, there is a significant gap in developing models that can handle multilingual, code-mixed, and dialect-rich student responses. Scalable ASAG solutions for India's broader linguistic landscape are still nascent.
9. **Yoga and Psychosomatic Health:** While ancient practices like yoga are validated for psychosomatic balance, there is insufficient empirical research linking specific yoga techniques with quantifiable improvements in psychosomatic illnesses within diverse Indian populations, considering cultural, gender, and socio-economic factors.
10. **Thermal Safety in Nanomaterials:** Research on the thermal stability of silica aerogels like MTMS is promising, yet broader application studies and comparative analyses with other insulating materials in diverse environmental conditions are lacking. The long-term sustainability and safety of these materials in real-world applications remain an open area.

IV. SOLUTIONS TO IDENTIFIED RESEARCH GAPS

1. **Multilingualism and Linguistic Diversity in India:** To address the paucity of empirical studies, a **longitudinal, cross-regional research framework** should be developed. This can integrate **sociolinguistic surveys, digital corpora creation, and ethnographic studies** in urban and rural educational settings. Utilizing **AI-based language tracking tools** can help map language use trends and shifts across generations, especially concerning English and regional languages in academic and professional domains.
2. **Language Preservation in Tribal Communities:** Develop a **technology-enabled revitalization model** combining **mobile apps, digital storytelling platforms, and interactive educational materials** in endangered languages like Gondi. Collaborations between linguists, policymakers, and tribal communities can establish **state-supported multilingual education policies and community-led language schools**, ensuring sustainable language transmission to younger generations.
3. **Emotion Regulation Measurement Tools in Indian Contexts:** Expand the validation of DERS-H by conducting **multisite psychometric studies** across different socio-economic, linguistic, and cultural groups in India. Adapting the tool with **regional language translations and cultural-specific emotion descriptors** will enhance its applicability. Additionally, developing **mobile-based emotion regulation assessment apps** can facilitate wider accessibility and usage.
4. **Machine Translation and Terminology for Low-Resource Languages:** Create **domain-specific, annotated parallel corpora** for various Indian languages with active participation from legal, medical, and technical experts. Leveraging **transfer learning and multilingual neural MT models** can improve terminology translation accuracy. Investment in **open-source MT research platforms** focused on Indian languages will democratize access to advanced translation technologies.
5. **Hindi ASR and Pooling Techniques Enhancement:** Research should pivot towards implementing **transformer-based architectures like Wav2Vec2 and Conformer** for Hindi ASR tasks. Further, develop **adaptive pooling methods** tailored for Hindi's phonetic diversity and noise conditions. Building **larger, diversified Hindi speech datasets** representing multiple dialects will improve model generalization and robustness.
6. **Cultural Representation in Hindi Cinema:** Introduce a **multilingual subtitling standard for Indian films**, ensuring accessibility across linguistic groups. Establishing **interdisciplinary research combining film studies, linguistics, and media psychology** can help assess the impact of linguistic representations on audience perceptions and cultural identity formation. Policies encouraging **regional language films in mainstream cinema spaces** should be supported.
7. **Gender, Entrepreneurship, and Technology:** Launch **regional women entrepreneurship incubation hubs** integrating ICT training, mentorship, and access to venture capital. Tailoring the co-innovation platform to address **region-specific socio-economic barriers** will increase participation. Partnerships with **local governments and international agencies** can ensure resource distribution and sustainability of these programs.
8. **Automatic Short Answer Grading in Low-Resource Languages:** Develop **deep learning models capable of handling code-mixed and dialect-rich responses**, with training on expanded datasets covering more Indian languages. Implementing **semi-supervised and transfer learning approaches** can compensate for data scarcity. Additionally, deploying these models in **educational platforms for real-time feedback** will aid in scaling the assessment process.
9. **Yoga and Psychosomatic Health:** Conduct **clinical trials integrating yoga with modern psychosomatic treatments**, segmented by demographic factors like age, gender, and socio-economic status. Develop **standardized yoga intervention protocols** for specific psychosomatic conditions. Collaborations between **medical researchers and traditional practitioners** will bridge scientific validation with traditional knowledge.
10. **Thermal Safety in Nanomaterials:** To enhance application scalability, conduct **comparative studies on silica aerogels versus other insulating materials** under varied environmental conditions. Develop **eco-friendly manufacturing techniques** to reduce production costs and environmental impact. Collaboration between **materials science researchers and industry stakeholders** will facilitate the commercialization of safer, more efficient thermal insulation products.

V. CONCLUSION AND FUTURE WORK

5.1 Conclusion : The exploration of machine translation (MT) and speech recognition for low-resource Indian languages reveals significant progress yet underscores persistent challenges. Current advancements, such as neural machine translation (NMT) and time-delay neural networks (TDNN) for speech recognition, have improved performance, but limitations remain due to data scarcity, dialectal variations, and domain-specific terminology gaps. Hindi and other regional languages still suffer from underrepresentation in linguistic datasets, impacting the accuracy and reliability of AI models. Moreover, code-mixing and sociolinguistic complexities inherent in Indian language usage further complicate model development and deployment. Existing approaches have demonstrated potential in clean and noisy environments, yet their generalization across languages and contexts is insufficient.

5.2 Future Work : Future research should focus on creating extensive, domain-specific parallel corpora across multiple Indian languages, incorporating legal, medical, and technical lexicons. Developing advanced models like transformers (e.g., Wav2Vec2, Conformer) tailored for speech recognition in diverse linguistic settings is essential. Efforts should also prioritize building datasets capturing code-mixed and dialect-rich speech for robust ASR systems. Integrating transfer learning, semi-supervised learning, and multilingual modeling can bridge data limitations. Collaborative frameworks between linguists, technologists, and policymakers will be crucial to

drive innovation, ensuring inclusive and scalable solutions for India's multilingual ecosystem.

References

1. Sridhar SN. Indian English. The handbook of Asian Englishes. 2020 Sep 1;241-77.
2. Zaki-Hindi A, Elayoubi SE, Chahed T. Multi-tenancy and URLLC on unlicensed spectrum: Performance and design. Computer Networks. 2020 Aug 4;177:107311.
3. Haque R, Hasanuzzaman M, Way A. Analysing terminology translation errors in statistical and neural machine translation. Machine Translation. 2020 Sep;34(2):149-95.
4. Mubarki MA. Body, masculinity and the male hero in Hindi cinema. Social Semiotics. 2020 Mar 14;30(2):225-53.
5. Iyer U. Bringing bharatanatyam to Bombay cinema: Mapping Tamil-Hindi film industry traffic through Vyjayanthimala's dancing body. In Industrial Networks and Cinemas of India 2020 Dec 13 (pp. 77-91). Routledge India.
6. Strauss S. Positioning yoga: Balancing acts across cultures. Routledge; 2020 Jun 3.
7. Boruah DM. Language Loss and Revitalization of Gondi language: An Endangered Language of Central India. language in India. 2020 Sep 1;20(9).
8. Bhatnagar P, Shukla M, Pandey R. Validating the factor structure of the Hindi version of the Difficulties in Emotion Regulation Scale. Journal of Psychopathology and Behavioral Assessment. 2020 Jun;42(2):377-96.
9. Siddiqui A, Siddiqui I. Usage of Urdu as the Language of Elitism among the Muslims of the Northern and the Deccan parts of India: A Socio-Cultural Review. Middle Eastern Journal of Research in Education and Social Sciences. 2020 Nov 3;1(2):107-15.
10. De Silva K, Basheer A, Antwi-Fordjour K, Beauregard MA, Chand V, Parshad RD. The "higher" status language does not always win: the fall of English in India and the rise of Hindi. Advances in Complex Systems. 2020 Dec 2;23(08):2050021.
11. Passricha V, Aggarwal RK. A comparative analysis of pooling strategies for convolutional neural network based Hindi ASR. Journal of Ambient Intelligence and Humanized Computing. 2020 Feb;11(2):675-91.
12. Sohail M. Origin and Development of Urdu Language in the SubContinent: Contribution of Early Sufia and Mushaikh. South Asian Studies. 2020 Jul 9;27(1).
13. Gupta C. Indentured women and Hindi Print-Public Sphere in early twentieth century India. In Women, gender and the legacy of slavery and indenture 2020 Nov 1 (pp. 101-120). Routledge.
14. Osmani MW, El Haddadeh R, Hindi N, Weerakkody V. The role of Co-innovation platform and E-collaboration ICTs in facilitating entrepreneurial ventures. International Journal of E-Entrepreneurship and Innovation (IJEI). 2020 Jul 1;10(2):62-75.
15. Kumar V, Mehta R. Adaptation and validation of receptive expressive emergent Language Test-3: Evidence from Hindi speaking children with cochlear implant. International Journal of Pediatric Otorhinolaryngology. 2020 May 1;132:109891.
16. Li Z, Zhang Y, Huang S, Wu X, Shi L, Liu Q. Thermal stability and pyrolysis characteristics of MTMS aerogels prepared in pure water. Journal of Nanoparticle Research. 2020 Oct;22(10):334.
17. Rahman M, Chen Y, Ibrahim A, Lindquist W, Tobias D, Krstulovich J, González D, Hindi R. Study of drying shrinkage mitigating concrete using scaled bridge bays. International Journal of Civil Engineering. 2020 Jan;18(1):65-73.
18. Agarwal D, Gupta S, Baghel N. ScAA: A dataset for automated short answer grading of children's free-text answers in Hindi and Marathi. In Proceedings of the 17th International Conference on Natural Language Processing (ICON) 2020 Dec (pp. 430-436).
19. Virk H, Seth S, Rabac JK, Jain D. Spirituality, yoga practice & natural dopamine effects. JS International Journal of Multidisciplinary Research. 2020 Dec 23;2(1).
20. Khatri JA. Favored Variety and Cultural Othering: Semantic Enclave in Hindi Cinema. Language in India. 2020 Mar 1;20(3).
21. Kumar A, Aggarwal RK. A time delay neural network acoustic modeling for Hindi speech recognition. In Advances in Data and Information Sciences: Proceedings of ICDIS 2019 2020 Jan 3 (pp. 425-432). Singapore: Springer Singapore.
22. Mubarki MA. Not on My Plate! Mapping the trajectory of the meat food culture of Hindi cinema. Social Semiotics. 2020 Mar 14;30(2):274-302.