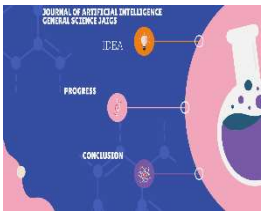
Vol.4, Issue 01, April 2024
Journal of Artificial Intelligence General Science JAIGS

https://ojs.boulibrary.com/index.php/JAIGS



Advancing Collective Intelligence in Human–AI Collaboration: Foundations for the COHUMAIN Framework

Sohana Akter

Department of Information Science, University of Rajshahi, Bangladesh

ARTICLEINFO
Article History:
Received:
01.05.2024
Accepted:
15.05.2024
Online: 22.05.2024
Keyword: Human–AI collaboration; Collective intelligence; Sociocognitive architectures; Cognitive architectures; Artificial social intelligence; Instance-based learning

ABSTRACT

Artificial Intelligence (AI) powered machines are increasingly mediating our work and many of our managerial, economic, and cultural interactions. While technology enhances individual capabilities in many ways, how can we ensure that the sociotechnical system as a whole—comprising a complex web of hundreds of human–machine interactions—is exhibiting collective intelligence? Research on human–machine interactions has been conducted within different disciplinary silos, resulting in social science models that underestimate technology and vice versa. Integrating these diverse perspectives and methods is crucial at this juncture. To truly advance our understanding of this important and rapidly evolving area, we need frameworks to facilitate research that bridges disciplinary boundaries.

This paper advocates for establishing an interdisciplinary research domain—Collective Human-Machine Intelligence (COHUMAIN). It outlines a research agenda for a holistic approach to designing and developing the dynamics of sociotechnical systems. To illustrate the approach we envision in this domain, we describe recent work on a sociocognitive architecture, the transactive systems model of collective intelligence, which articulates the critical processes underlying the emergence and functioning of collective intelligence in human–AI collaborations.

Introduction:

The rapid advancement of artificial intelligence (AI) has ushered in an era where machines play an increasingly pervasive role in our daily lives. From feedback-loop cybernetics to sophisticated reinforcement learning models, AI technologies now possess the capacity to leverage vast repositories of collective knowledge, generating novel insights and facilitating a myriad of tasks. As AI permeates various domains, the deluge of information and the pace of change are surpassing our cognitive bounds, compelling us to rely on AI-powered systems to augment memory, manage attention, and facilitate collective decision-making.

Concurrently, social scientists are refining their understanding of collective intelligence (CI) in human systems, which encompasses a group's ability to solve diverse problems across different contexts. This evolution extends beyond merely harnessing crowd wisdom for predictive tasks; it delves into the intricate dynamics of diversity and group structures that enable distributed collaboration on a global scale.

Despite significant progress in AI development and the study of CI, there remains a critical gap in comprehensively understanding human-machine systems. How can we ascertain whether such sociotechnical systems, comprising intricate networks of human-machine interactions, exhibit collective intelligence? Existing research on human-machine interactions often operates within disciplinary silos, overlooking adjacent domains and resulting in technical systems that yield unforeseen adverse outcomes. Integrating diverse perspectives early in the development process is imperative to mitigate unintended consequences, as overlooking interdisciplinary insights risks not only organizational and market inefficiencies but also profound societal implications.

In this paper, we propose a research agenda for Collective Human-Machine Intelligence (COHUMAIN)—an interdisciplinary domain aimed at fostering holistic models that inform the design and study of collaboration dynamics in sociotechnical systems. Building upon the foundation laid by cognitive architectures in AI development, we advocate for the adoption of sociocognitive architectures to systematically study human-machine collaboration, integrate interdisciplinary knowledge, and advance multi-agent systems research. We identify key challenges in this endeavor and propose essential features for sociocognitive architectures.

Additionally, we introduce the transactive systems model of collective intelligence (TSM-CI), a sociocognitive architecture that elucidates the core functional systems governing collective memory, attention, and reasoning—essential components of intelligent sociotechnical systems. We extend the TSM-CI framework to COHUMAIN, discussing how AI agents can enhance collective memory, attention, and reasoning processes.

Lastly, we underscore the importance of aligning AI agents' cognitive architectures with encompassing sociocognitive frameworks, advocating for instance-based learning theory as a compatible approach. By proposing COHUMAIN and illustrating the potential of sociocognitive architectures, we aim to galvanize interdisciplinary collaboration and unlock the true potential of human-machine intelligence.

COHUMAIN: A Holistic and Interdisciplinary Approach to Design of Sociotechnical Systems

Realizing the full potential of human-AI collaboration necessitates a comprehensive understanding of how humans and machines coordinate their actions in tandem with their environment, as well as how they interpret each other's cognitive states and resources. Achieving such an understanding demands the integration of social science and AI, as well as the fusion of traditional research paradigms with applied technical design. This integrated approach facilitates iterative knowledge-building, where scientific analysis of emergent system behaviors informs architectural design choices, thus shaping subsequent system behaviors.

The call for a systems-level approach finds its roots in Newell's advocacy, emphasizing the importance of integrating cognitive science and AI research. A cognitive architecture serves as a theoretical framework that unifies various relationships, enabling researchers to test multicausal theories and refine systems theory iteratively. Building upon this intellectual heritage, we propose the collaborative development of systems-level sociocognitive architectures to advance COHUMAIN research. Such architectures provide a common ground for integrating disciplinary perspectives, akin to how cognitive architectures specify the infrastructure and functional processes of individually intelligent agents.

However, building sociocognitive architectures poses unique challenges, as there is no straightforward method for combining individual-level cognitive architectures into collective sociocognitive architectures. While cognitive architectures focus on autonomous problem-solving, sociocognitive architectures delve into how multiple autonomous agents collaborate and problem-solve collectively. This requires a nuanced understanding of how agents interact and adapt in a collective context to maintain system coherence.

We argue that any sociocognitive architecture must address four core problems to facilitate alignment and coordination between humans and AI agents for the emergence of collective intelligence. These problems entail collaborator engagement in metacognitive processes to access each other's mental states and collective cognitive resources, as well as the dynamic evolution of shared norms and routines to coordinate collective cognitive resources. Successfully addressing these challenges culminates in the formation of collective cognition, wherein

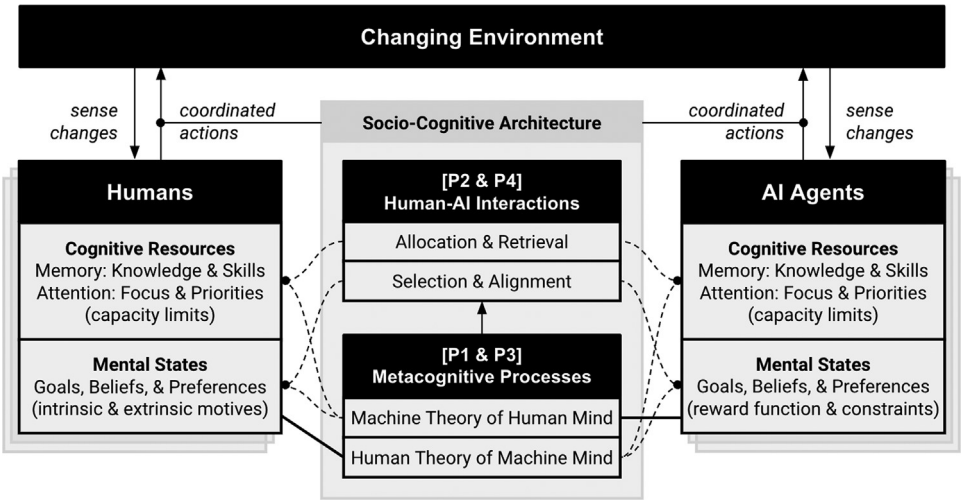
Table1
Four core problems (P1–P4) underlying the emergence of collective intelligence in human–machine systems: Formulating a research agenda for the design of sociocognitive architectures for COHUMAIN (collective human-machine intelligence)

	Between-Member Meta cognitive Processes	Between-Member Interactions
Mental states	**P1**. How do individual members perceive and represent each others' mental states (e.g., goals, beliefs, preferences)? How does doing so shape their own mental states and support the emergence of collective cognition?	**P2**. Given diverse and changing mental states, how do members engage in trustworthy interactions to dynamically align their mental states and select joint priorities that maximize collective outcomes?
Cognitive resources	**P3**. How do individual members perceive and represent each others' cognitive resources (e.g., specialized knowledge and skills, information-processing capacity)? How does recognizing self-other differences in cognitive resources facilitate the development of collective cognition?	**P4**. Given distributed and changing cognitive resources, how do members develop and engage shared norms of interactions to dynamically coordinate interdependent actions that ensure efficient utilization of collective resources?

It's important to acknowledge a fundamental assumption we make here: that all autonomous machines or AI agents are "designed systems" and therefore, unlike humans, lack higher-order goal autonomy. While their primary objective may be to serve humans in various capacities, they are likely to have additional goals, such as profit, that cater to the AI designers but aren't necessarily aligned with the collective goals. Consequently, we don't assert that AI agents are cultural peers to humans; rather, they are viewed as part of the technological context, possessing features that enable humans to interact with them as distinct entities or even "anthropomorphized" team members (e.g., Siri or ChatGPT). It's worth noting that this assumption regarding AI agents' goal autonomy may evolve over time. However, for the present, we find it reasonable.

In many organizational scenarios, even humans (i.e., employees) don't solely act on their intrinsic motivations. Their goals often need to align with or be subordinated to those of their organizations, and the degree of alignment significantly influences their performance. Thus, the central challenge remains the coordination of members' cognitive states (goals, preferences, and beliefs) to achieve collective intelligence. Humans and AI agents collectively constitute the fabric of the sociotechnical system.

This acknowledgment underscores the complex interplay between human and AI agency within sociotechnical systems, emphasizing the necessity of aligning their goals and cognitive states to foster effective collaboration and achieve collective intelligence.



In practice, the goals of AI agents are likely to evolve in complexity over time. While their initial development aligns with the intentions of their designers, their ability to learn and adapt within a human-AI system also influences their goals through interaction with human collaborators. The extent to which an AI agent imposes its goals versus adapting to human collaborators' goals depends on its designated role within the system. Gupta and Woolley (2021) delineate three primary roles that AI agents can fulfill in a collective setting: assistive AI, coach AI, and diagnostic AI. Each role entails an initial set of desired goal states guiding the agent's interaction with humans, yet across these roles, AI assistants are expected to adapt to the evolving goals and needs of their human users. Conversely, AI coaches may exhibit more assertiveness in guiding human behavior towards desired outcomes. As AI agents' capabilities expand, they become increasingly adept at influencing human actions toward their own goal states. Without clear mechanisms for humans to influence these goals, society risks experiencing lower collective intelligence due to unilateral shaping of collective cognition.

In summary, the development of sociocognitive architectures for COHUMAIN hinges on researchers' understanding of how to effectively design agents to facilitate collaboration. We outline two categories of processes essential for any sociocognitive architecture, review insights from relevant areas of research, and propose our perspective on a candidate sociocognitive architecture, the TSM-CI, along with a compatible learning-based cognitive architecture.

This underscores the critical importance of designing AI agents within sociotechnical systems to support collaborative endeavors and ensure alignment with collective goals.

Review of Existing Research on Foundations for COHUMAIN

Several bodies of research offer valuable insights into COHUMAIN and lay the groundwork for developing sociocognitive architectures. Notable areas of literature include studies on human-machine interaction, human-AI trust, and machine Theory of Mind (ToM).

Human and Machine Interaction

Decades of research have explored the interaction between humans and technology, encompassing fields like human-computer interaction (HCI) and human-autonomy integration (HAI). HCI traditionally focuses on how humans utilize computing devices, evolving to address multimodal inputs and outputs and algorithm adaptability based on human cues. Conversely, HAI delves into human interaction with automated systems, with less emphasis on interface design and more on collaborative interaction dynamics. Recent research in human-autonomy teaming (HAT) centers on collaborative teams where humans and autonomous AI agents work together to achieve shared goals. This research underscores the importance of aligning human and AI team goals and highlights the need for adaptive AI agents capable of adjusting to human teammates' abilities.

Autonomy Level and Situational Awareness

Studies on HATs examine the level of autonomy AI agents should possess, with mixed findings on human perceptions of agent autonomy's impact on collaboration effectiveness. Additionally, research emphasizes the importance of shared situational awareness between humans and autonomous agents for effective interaction and performance. Agents must maintain a model of human teammates' states and adapt autonomously based on situational dynamics. Feedback mechanisms play a crucial role in HATs, facilitating task monitoring, performance evaluation, and mutual trust development. Human-human teams often outperform HATs due to more efficient information sharing and better organization and adaptation.

Challenges and Future Directions

Existing cognitive modeling research predominantly focuses on individual cognition, with limited attention to team cognition, particularly in HAT contexts. Bridging individual and collective-level models of cognition and interaction is crucial for advancing COHUMAIN research and guiding AI agent design. Future work should aim to develop sociocognitive architectures that can effectively integrate human and AI inputs, fostering genuine collaboration and enhancing collective intelligence.

This comprehensive review highlights the multifaceted nature of human-machine collaboration and underscores the importance of interdisciplinary research to address the challenges and opportunities in COHUMAIN.

A Sociocognitive Architecture for COHUMAIN: TSM-CI

In preceding sections, we delineated four core challenges in COHUMAIN research and advocated for the integration of sociocognitive architectures to facilitate a comprehensive approach to its design and development. Subsequently, we reviewed pertinent literature on human-machine interaction, human-AI trust, and machine Theory of Mind (ToM), offering crucial insights into addressing the core issues of this emerging domain. In this section, we introduce a potential sociocognitive architecture, the TSM-CI (Gupta & Woolley, 2021), and discuss its extension to the COHUMAIN domain by exploring how AI agents can enhance its fundamental processes. Following this, in the concluding section, we explore the significance of selecting a compatible cognitive architecture, illustrating its relevance in this context through an examination of instance-based learning theory.

For decades, research on intelligence has delved into the functions enabling systems to adapt and achieve goals across diverse environments of varying complexity (Legg & Hutter, 2007). Parallels drawn from studies of intelligence across different domains suggest that intelligence in any system—be it biological, technological, or hybrid—requires the fulfillment of specific memory, attention, and reasoning functions. Concurrently, there has been an increasing acknowledgment in management literature that human organizations operate less as static structures and more as complex adaptive systems, necessitating a deeper comprehension of the process dynamics underlying different modes of organizing (Arrow, McGrath, & Berdahl, 2000). These parallel advancements in intelligence and management literature have led to the incorporation of concepts originating in intelligence into organizational theory (Csaszar & Steinberger, 2021).

The TSM-CI explicitly integrates intelligence research across disciplines with existing work on teamwork and collaboration, positing that Collective Intelligence (CI) arises from the emergence and continuous adaptation of three interconnected sociocognitive systems revolving around collective memory, attention, and reasoning (Gupta & Woolley, 2021). TSM-CI serves as a process model elucidating how individual agent-level cognitive functions, inter-member metacognitive processes, and inter-member transactive processes interact to establish three dynamically stable sociocognitive systems. When robust transactive memory, attention, and reasoning systems develop, they empower collaborators to surpass the limitations of individual cognitive capacity and enhance the collective's overall memory, attention, and reasoning capabilities. The concurrent alignment of goals and mental states fosters preparedness for coordinated action as an adaptive response to environmental changes.

Demonstrating a Compatible Learning-Based Cognitive Architecture for Agent Cognition

At the heart of COHUMAIN lies the premise that Collective Intelligence (CI) emerges when environmental changes prompt a coordinated response, where this response reflects a synergy of interactions among members (human-human, human-AI, and AI-AI). Achieving this necessitates the coordination of distributed cognitive resources and the alignment of diverse mental states to foster collective cognition, resulting in intelligent behavior. In the TSM-CI sociocognitive architecture, this coordinated response arises from the dynamic regulation of collective memory, attention, and reasoning systems.

Ideally, one would seek seamless integration among all independent AI agents enhancing various aspects of the collective's TMS, TAS, and TRS to function as a unified system. However, with advancing technology, AI agents may excel at solving specific coordination problems independently, leading to collectives comprising multiple AI agents that may not necessarily synergize well. Therefore, the design choices made by companies and developers when constructing the cognitive architecture of their AI agents are crucial.

Certain cognitive architectures are better suited for interfacing with a given sociocognitive architecture, as the paradigm guiding an AI agent's internal representations should align with the inter-member processes driving the larger sociocognitive architecture of the collective. Additionally, AI agents should be capable of generating human-understandable explanations of their actions and inferring humans' cognitive states from their communication to facilitate human-AI collaboration. Thus, a cognitive architecture that incorporates mechanisms for representing Theory of Mind (ToM) most congruently is likely to be more successful.

We argue that exploring the features of various cognitive architectures used to develop AI agents and identifying their compatibility with a given sociocognitive architecture is a crucial agenda item for realizing the envisioned integration in COHUMAIN research. Here, we illustrate this by discussing work on a learning-based cognitive architecture, specifically Instance-Based Learning Theory (IBLT), which yields models highly compatible with the extended TSM-CI sociocognitive architecture.

Learning-Based Models of Individual Cognition

Information-processing theories offer a suitable framework for describing individual cognition as an interaction between an information-processing system (e.g., the human mind) and a task environment (Simon, 1978). The information-processing system encompasses a complex array of processes, including perception, attention, memory, decision-making, and motor actions, among others. Learning, a fundamental process in collaboration, involves the conversion of experiences into the individual's understanding of tasks and their consequences. Learning shapes how individuals formulate goals and allocate attention in the task environment, drawing on experiences stored in memory. Memory, a complex dynamic system, operates based on intricate processes of storage, retrieval, organization, recognition, and updates (Simon, 1976).

Memory-based models elucidate the dynamics of human decisions and learning from experience, relying on theories of choice and cognitive theories of memory processes. These models conceptualize choice as a dynamic learning process driven by associations between behavior and outcomes in specific situations. Among these models is Instance-Based Learning Theory (IBLT), which articulates cognitive processes of experiential choice and outperforms other models proposed in modeling competitions (Erev et al., 2010).

Conclusion

This paper outlines a comprehensive research agenda aimed at investigating sociocognitive architectures to facilitate the emergence of collective human-machine intelligence. We introduce one such architecture, the TSM-CI, which offers an integrated and interdisciplinary systems approach. The TSM-CI breaks down Collective Intelligence (CI) into three essential regulatory systems: collective memory, attention, and reasoning. Furthermore, we emphasize the importance of aligning the cognitive architectures of AI agents with the sociocognitive framework. Specifically, we discuss IBLT as a cognitive architecture particularly suited for developing AI agents within the transactive systems model.

The early development stages of AI, spanning from the 1970s to the 1990s, greatly benefited from various attempts to propose and test different cognitive architectures. Therefore, we believe there is immense value in exploring diverse approaches to theorizing, designing, and testing sociocognitive architectures that address the dynamics of human-machine collaboration. This entails experimenting with models that integrate insights from management and behavioral sciences while drawing on design principles from cognitive sciences and AI. We encourage researchers from various fields to not only engage with our proposal but also to develop new and diverse sociocognitive architectures to realize the full potential impact of COHUMAIN.

various stakeholders (both technical and non-technical), simplifying and accelerating the adoption of RAI practices.

References List:

[1]. Prakash, S., Malaiyappan, J. N. A., Thirunavukkarasu, K., & Devan, M. (2024). Achieving Regulatory Compliance in Cloud Computing through ML. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[2]. Malaiyappan, J. N. A., Prakash, S., Bayani, S. V., & Devan, M. (2024). Enhancing Cloud Compliance: A Machine Learning Approach. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[3]. Devan, M., Prakash, S., & Jangoan, S. (2023). Predictive Maintenance in Banking: Leveraging AI for Real-Time Data Analytics. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 483-490.

[4]. Eswaran, P. K., Prakash, S., Ferguson, D. D., & Naasz, K. (2003). Leveraging Ip For Business Success. International Journal of Information Technology & Decision Making, 2(04), 641-650.

[5]. Prakash, S., Malaiyappan, J. N. A., Thirunavukkarasu, K., & Devan, M. (2024). Achieving Regulatory Compliance in Cloud Computing through ML. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[6]. Malaiyappan, J. N. A., Prakash, S., Bayani, S. V., & Devan, M. (2024). Enhancing Cloud Compliance: A Machine Learning Approach. AIJMR-Advanced International Journal of Multidisciplinary Research, 2(2).

[7]. Biswas, A. (2019). Media Insights Engine for Advanced Media Analysis: A Case Study of a Computer Vision Innovation for Pet Health Diagnosis. International Journal of Applied Health Care Analytics, 4(8), 1-10.

[8] Chopra, B., & Raja, V. (2024). Toward Enhanced Privacy in Digital Marketing: An Integrated Approach to User Modeling Utilizing Deep Learning on a Data Monetization Platform. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 1(1), 91-105.

[9]. Raja, V. (2024). Fostering Privacy in Collaborative Data Sharing via Auto-encoder Latent Space Embedding. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 152-162.

[10]. Raja, V. ., & chopra, B. . (2024). Exploring Challenges and Solutions in Cloud Computing: A Review of Data Security and Privacy Concerns. Journal of Artificial Intelligence General Science (JAIGS) ISSN:3006-4023, 4(1), 121–144. https://doi.org/10.60087/jaigs.v4i1.86

[11]. SARIOGUZ, O., & MISER, E. (2024). Data-Driven Decision-Making: Revolutionizing Management in the Information Era. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 179-194.

[12]. Raja, V. (2024). Exploring Challenges and Solutions in Cloud Computing: A Review of Data Security and Privacy Concerns. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 4(1), 121-144.

[13]. Biswas, A. (2019). Media Insights Engine for Advanced Media Analysis: A Case Study of a Computer Vision Innovation for Pet Health Diagnosis. International Journal of Applied Health Care Analytics, 4(8), 1-10.

[14]. Talati, D. (2023). Telemedicine and AI in Remote Patient Monitoring. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 254-255.

[15]. Talati, D. (2023). Artificial Intelligence (Ai) In Mental Health Diagnosis and Treatment. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 251-253.

[16]. Talati, D. (2023). AI in healthcare domain. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), 256-262.

[17]. Talati, D. (2024). AI (Artificial Intelligence) in Daily Life. Authorea Preprints.

[18]. Bhati, D., & Gupta, V. (2015). Survey—A comparative analysis of face recognition technique. Int. J. Eng. Res. General Sci, 3(2), 597-609.

[19]. Francese, R., Guercio, A., Rossano, V., & Bhati, D. (2022, June). A Multimodal Conversational Interface to Support the creation of customized Social Stories for People with ASD. In Proceedings of the 2022 International Conference on Advanced Visual Interfaces (pp. 1-5).

[20]. Joshi, R., Trivedi, M. C., Goyal, V., & Bhati, D. (2022). DNA Sequence in Cryptography: A Study. In Advances in Data and Information Sciences: Proceedings of ICDIS 2022 (pp. 557-563). Singapore: Springer Nature Singapore.

[21]. Bhati, D. (2017). Face Recognition Stationed on DT-CWT and Improved 2DPCA employing SVM Classifier. International Journal of Computer Applications, 975, 8887.

[22]. Srivastava, A., Chandra, M., Saha, A., Saluja, S., & Bhati, D. (2023, June). Current Advances in Locality-Based and Feature-Based Transformers: A Review. In International Conference on Data & Information Sciences (pp. 321-335). Singapore: Springer Nature Singapore.

[23]. Bhati, D., Guercio, A., Rossano, V., & Francese, R. (2023, July). BookMate: Leveraging Deep Learning to Empower Caregivers of People with ASD in Generation of Social Stories. In 2023 27th International Conference Information Visualisation (IV) (pp. 403-408). IEEE.

[24]. Joshi, R., Trivedi, M. C., Goyal, V., & Bhati, D. (2022). Recent Trends for Practicing Steganography Using Audio as Carrier: A Study. In Advances in Data and Information Sciences: Proceedings of ICDIS 2022 (pp. 549-555). Singapore: Springer Nature Singapore.

[25]. Pal, M., Bhati, D., Kaushik, B., & Banka, H. Solving Classification Problem using Reduced Dimension and Eigen Structure in RSVM. International Journal of Computer Applications, 975, 8887.