Research Statement

Nagur Shareef Shaik, shaiknagurshareef6@gmail.com
Master of Science in Computer Science, Georgia State University, Atlanta
Google Scholar: https://scholar.google.com/citations?user=rkfPEjUAAAAJ

My journey into Machine Learning began with a curiosity sparked by a simple question: "How can machines learn to understand and interpret complex data like humans do?" What started as fascination soon transformed into a deep-seated passion as I delved deeper into the realms of computer vision, natural language processing, and their intersection with biomedical sciences. Imagine starting with lines of code and ending with systems that can analyze images and text to diagnose diseases and generate insightful clinical reports. The ability to harness the power of data, to teach machines not just to see pixels and characters but to understand patterns and meanings, fueled my passion further and became my driving force for a relentless pursuit to develop intelligent systems. The journey was not without its challenges. From grappling with the complexities of multi-modal data integration to fine-tuning the intricacies of attention mechanisms that guide where the model should focus its learning, every hurdle presented an opportunity to innovate and refine. My focus on designing robust multi-modal fusion strategies became paramount in transforming multi-modal data into actionable intelligence.

Currently, I'm deeply immersed in designing and training large language models (LLMs) tailored for precise insights in healthcare. Progressing further, I envision pioneering vision language models (VLMs) that revolutionize diagnostic and decision support systems by analyzing medical images and interpreting them within real-world contexts. Enhancing the Contrastive Language-Image Pre-training (CLIP) framework is crucial, optimizing its ability to learn from images and text through advanced contrastive methods. I'm also refining a robust strategy to seamlessly integrate insights from multi-modal data, customizing LLMs to decode integrated representations effectively. My goal is to deliver a high-performance VLM with optimized parameters, empowering professionals with advanced tools for extracting nuanced insights from combined visual and textual data.

In the dynamic field of medical AI, my journey has been marked by significant strides across various domains. Recently, our focus on Vision Language Models (VLMs) reached a pinnacle with the development of an Optimal VLM. This cutting-edge model integrates advanced features like Guided Context Gating Attention for robust vision encoding and Transformer-based decoding, enabling precise generation of clinical reports. In Brain Tumor Recognition, innovations such as Multi-modal Squeeze and Excitation attention enhanced the detection of tumor-specific features, while a subsequent multi-level attention network improved precision by 6%. Our approach to Diabetic Retinopathy Severity Classification began with a foundational CNN-based framework, evolving through gated attention models and a hybrid Neural Support Vector Machine, resulting in a remarkable 9% boost in diagnostic accuracy. Innovations in COVID-19 Detection included pioneering an ensemble method based on prediction confidence values, enhancing model resilience and accuracy. Moving forward, my research is dedicated to extending these breakthroughs to 3D medical imaging modalities like sMRI and fMRI, aiming to develop intelligent diagnostic systems despite challenges with annotated data. These endeavors underscore my commitment to advancing AI-driven solutions that harness the power of multi-modal data integration to improve healthcare outcomes.

These research efforts have yielded significant contributions, with methods and results from innovations including Spatial Sequence Attention, Guided Context Gating, and M3 Transformer published in 14 reputed journals and presented at 3 top-tier conferences, including the International Symposium on Biomedical Imaging (ISBI) and the International Conference on Image Processing (ICIP). These presentations highlight the advancements made in integrating advanced attention mechanisms and transformative models in the field of medical AI, contributing to the forefront of research at these esteemed venues.

My research focuses on Natural Language Processing (NLP), Computer Vision (CV), Multimodal Data Handling, and Model Optimization and Efficiency, aligning with the transformative role of AI in medicine. Techniques such as fine-tuning large language models (LLMs), in-context learning, and multi-modal fusion are pivotal for advancing clinical decision-making, diagnostics, and personalized care. These approaches enable the integration of diverse data—such as medical reports, imaging data, and patient history—into unified representations that capture richer contextual information. For instance, combining radiology images with clinical notes allows for more accurate diagnosis and treatment recommendations. Multi-modal fusion techniques further enhance disease detection, prognosis, and patient outcome prediction by efficiently organizing and analyzing comprehensive medical data, facilitating precise and timely interventions.

Research Publications

[1] T. K. Cherukuri, N. S. Shaik, and D. H. Ye, "Guided context gating: Learning to leverage salient lesions in retinal fundus images," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Paper ID: 1605, 2024.

[2] N. S. Shaik and T. K. Cherukuri, "Gated contextual transformer network for multi-modal retinal image clinical description generation," *Image and Vision Computing*, vol. 143, no. C, 2024.

[3] N. S. Shaik, T. K. Cherukuri, V. Calhoun, and D. H. Ye, "Spatial sequence attention network for schizophrenia classification from structural brain mr images," in *Proceedings of the 21st IEEE International Symposium on Biomedical Imaging*, IEEE, 2024.

[4] N. S. Shaik, T. K. Cherukuri, N. Veeranjaneulu, and J. D. Bodapati, "Medtransnet: Advanced gating transformer network for medical image classification," *Machine Vision and Applications*, vol. 35, no. 4, p. 73, 2024.

[5] N. S. Shaik, T. K. Cherukuri, and D. H. Ye, "M3t: Multi-modal medical transformer to bridge clinical context with visual insights for retinal image medical description generation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Paper ID: 1604, 2024.

[6] N. S. Shaik and T. K. Cherukuri, "Visual attention based composite dense neural network for facial expression recognition," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 12, pp. 16 229–16 242, 2023.

[7] N. S. Shaik and T. K. Cherukuri, "Hinge attention network: A joint model for diabetic retinopathy severity grading," *Applied Intelligence*, vol. 52, no. 13, pp. 15 105–15 121, 2022.

[8] N. S. Shaik and T. K. Cherukuri, "Multi-level attention network: Application to brain tumor classification," *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 817–824, 2022.

[9] N. S. Shaik and T. K. Cherukuri, "Transfer learning based novel ensemble classifier for covid-19 detection from chest ct-scans," *Computers in Biology and Medicine*, vol. 141, p. 105 127, 2022.

[10] J. D. Bodapati, N. S. Shaik, and V. Naralasetti, "Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 10, pp. 9825–9839, 2021.

[11] J. D. Bodapati, N. S. Shaik, and V. Naralasetti, "Deep convolution feature aggregation: An application to diabetic retinopathy severity level prediction," *Signal, Image and Video Processing*, vol. 15, pp. 923–930, 2021.

[12] J. D. Bodapati, N. S. Shaik, V. Naralasetti, and N. B. Mundukur, "Joint training of two-channel deep neural network for brain tumor classification," *Signal, Image and Video Processing*, vol. 15, no. 4, pp. 753–760, 2021.

[13] J. D. Bodapati, S. N. Shareef, V. Naralasetti, and N. B. Mundukur, "Msenet: Multi-modal squeeze-and-excitation network for brain tumor severity prediction," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 35, no. 07, p. 2 157 005, 2021.

[14] N. S. Shaik and T. K. Cherukuri, "Lesion-aware attention with neural support vector machine for retinopathy diagnosis," *Machine Vision and Applications*, vol. 32, no. 6, p. 126, 2021.

[15] J. D. Bodapati, V. Naralasetti, S. N. Shareef, S. Hakak, M. Bilal, P. K. R. Maddikunta, and O. Jo, "Blended multi-modal deep convnet features for diabetic retinopathy severity prediction," *Electronics*, vol. 9, no. 6, p. 914, 2020.

[16] V. Dondeti, J. D. Bodapati, S. N. Shareef, and N. Veeranjaneyulu, "Deep convolution features in non-linear embedding space for fundus image classification.," *Rev. d'Intelligence Artif.*, vol. 34, no. 3, pp. 307–313, 2020.

[17] J. D. Bodapati, N. Veeranjaneyulu, and S. N. Shareef, "Sentiment analysis from movie reviews using lstms.," *Ingénierie des Systèmes d Inf.*, vol. 24, no. 1, pp. 125–129, 2019.