Confidence Detection System

Real-time Multimodal Confidence Analysis

RUBAYET ALAM AZAN (22-46888-1), S.M. NAFEES HOSSAIN NILOY (22-47333-2), AL FAHAD (22-47125-1), ABDULLAH AL MARJUC TALHA (22-47294-1)

Introduction

The rise of virtual communication and remote interaction has led to a greater necessity for immediate confidence evaluation. Regular self-reports are not very useful because they are not immediate and not objective. This study combines the analysis of facial expressions, recognition of hand gestures, and vocal analysis to produce a dynamic confidence score. Real-time systems that utilize multiple cues are more accurate and provide better monitoring of engagement and adaptive support in digital environments. The system takes advantage of recent developments in machine learning, integrating different modes of communication, and real-time processing to provide uninterrupted and trustworthy feedback.

Methodology

This study follows the Design Science Research Methodology (DSRM), beginning with the identification of the core problem: the difficulty of assessing user confidence in virtual environments where traditional cues are harder to interpret. To address this, the system was designed using MediaPipe Face Mesh with 468 facial landmarks, MediaPipe Hands with 21 hand landmarks, the Web Audio API for real-time vocal analysis, and JavaScript to enable lightweight, browser-based processing. Feature extraction focused on capturing facial behaviors such as gaze direction, blink rate, head pose, expressions, and lip movement; hand-related behaviors including gesture clarity, speed, palm orientation, and fidgeting; and vocal characteristics encompassing pitch, jitter, loudness, and spectral centroid. These multimodal features were combined to compute a unified confidence score, where vocal features contributed 40%, facial features 30%, and hand gestures 30%, ensuring a balanced weighted fusion that generated scores between 0 and 100. The system was then tested and optimized through simulated online interviews and classroom scenarios, comparison with human evaluators, and targeted latency improvements to ensure smooth real-time performance even on low-end devices.

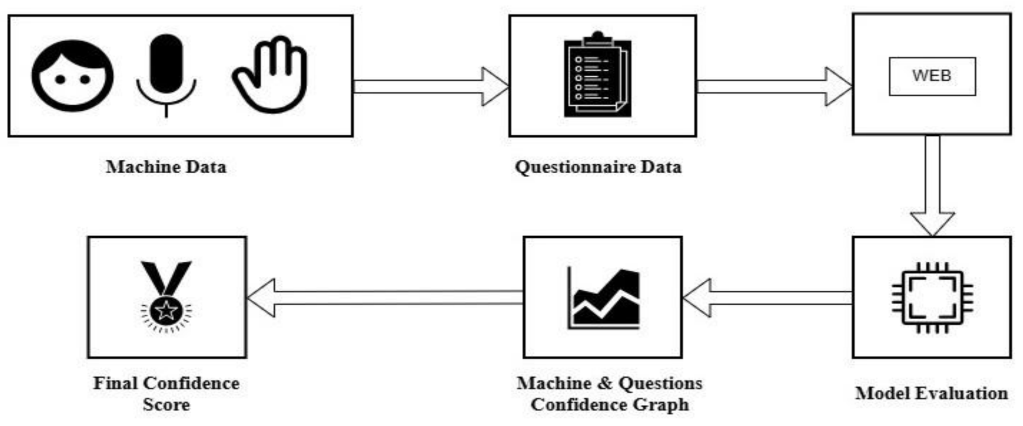


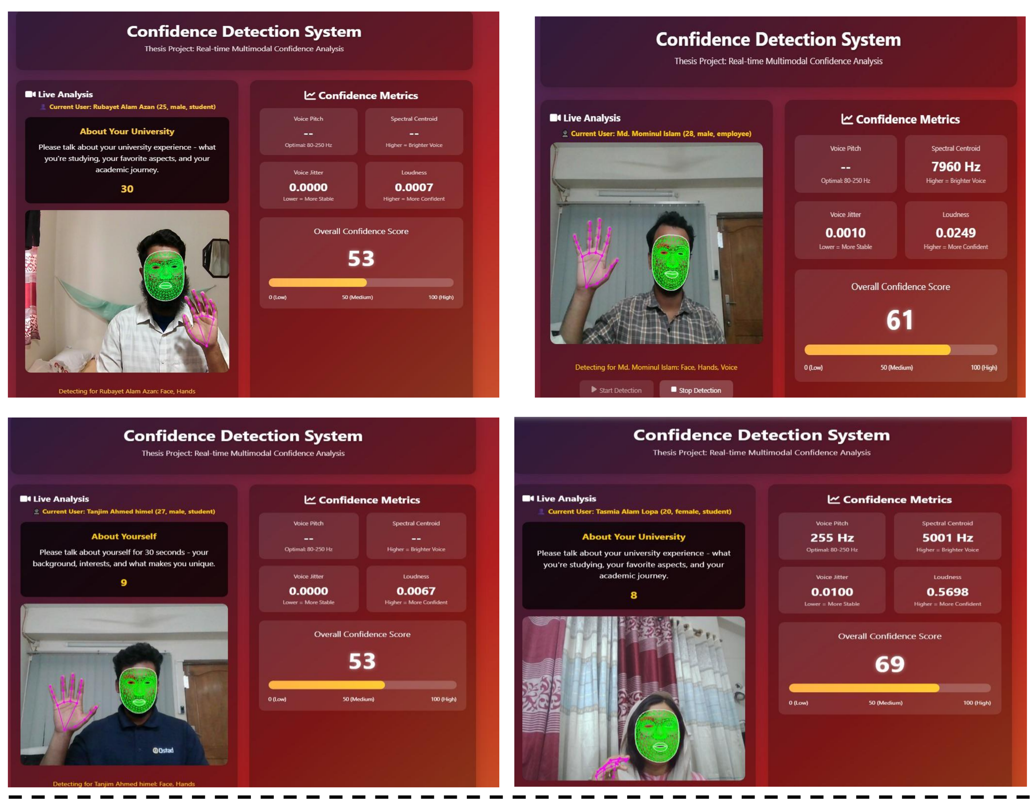
Fig 1. A flowchart illustrating the proposed model for real-time image processing and confidence estimation.

Data Analysis

A. **Machine Data Analysis:** The researchers studied facial features along with hand, vocal characteristics, to find patterns which indicate confidence levels. The main indicators consisted of smile intensity together with blink rate, head movement, and gesture stability, and pitch consistency, and loudness balance. The evaluation process used these non-verbal indicators to assess participant focus and their instantaneous self-assurance levels.

B. **Questionnaire Data Analysis**: The participants finished a pre-designed confidence questionnaire which asked them to rate their own feelings through self-assessment. The collected answers served as a benchmark to evaluate the accuracy of machine-generated confidence scores by verifying their consistency.

C. **Combined Confidence Evaluation**: The final confidence graph and score emerged from the combination of machine data with questionnaire results. The integrated method improved accuracy through its ability to verify human observation results against system prediction outcomes.



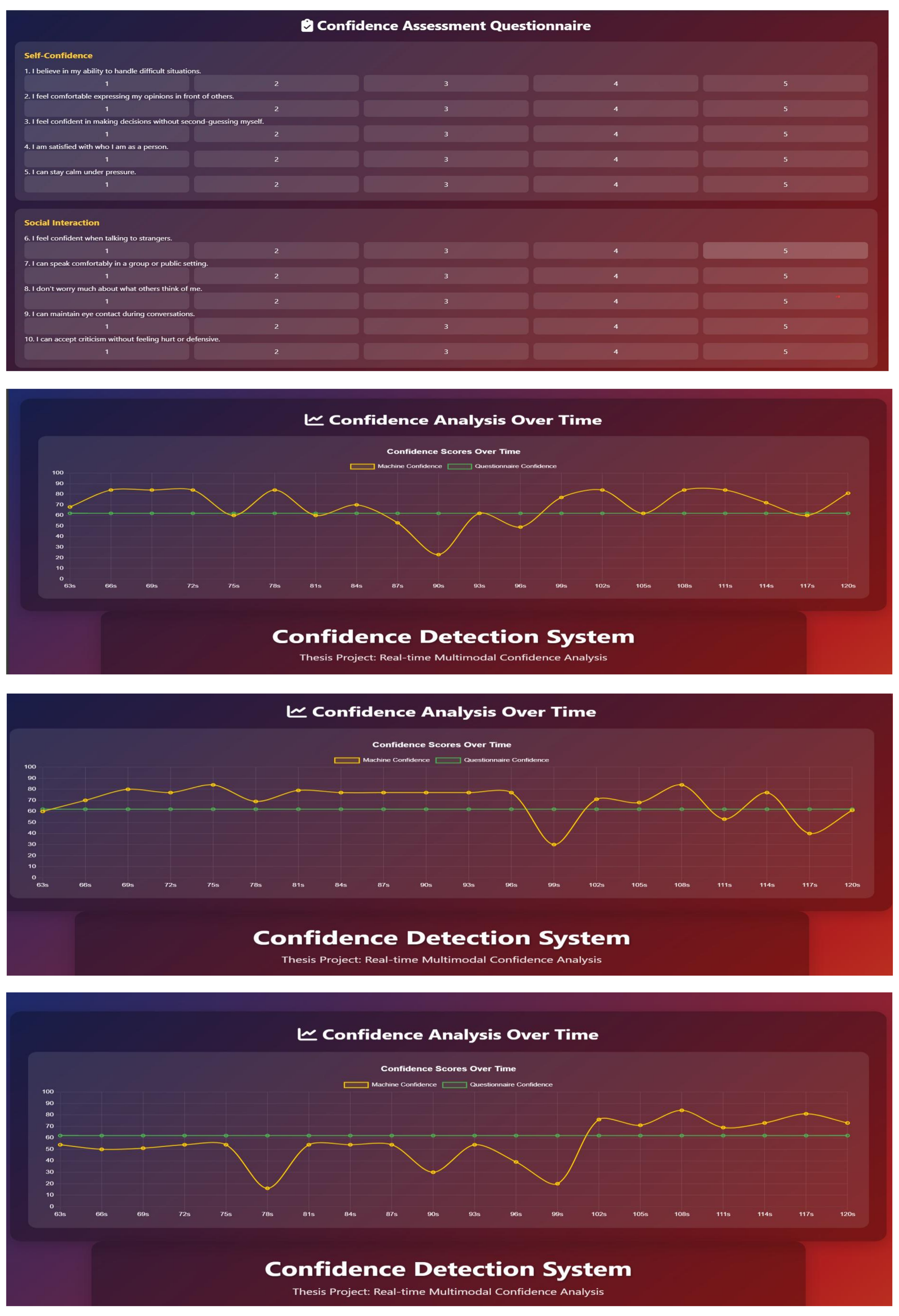
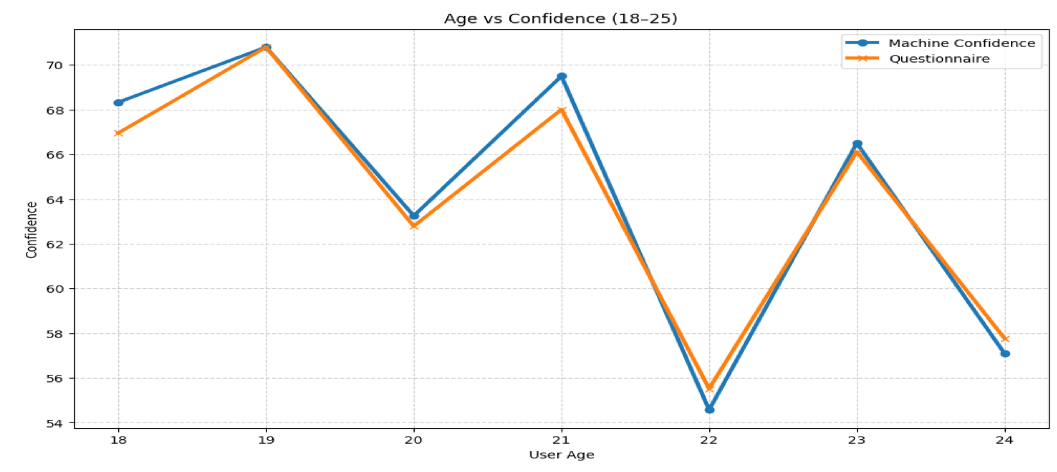
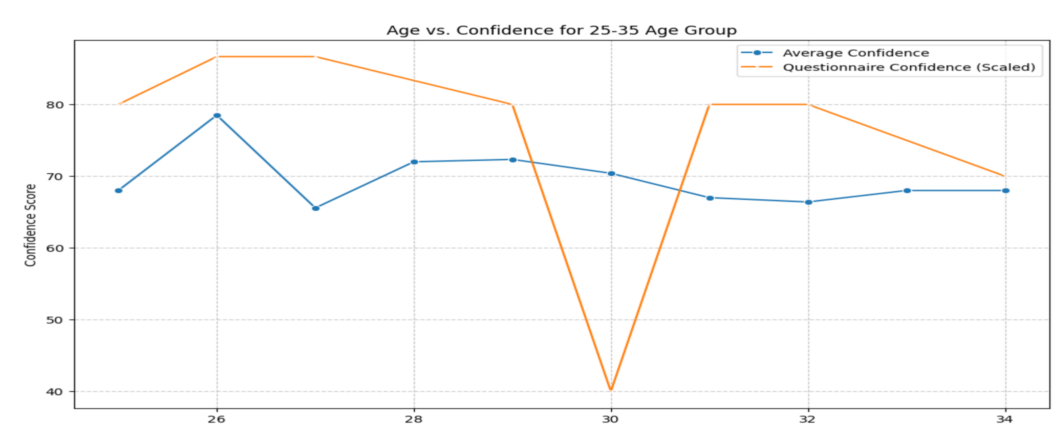


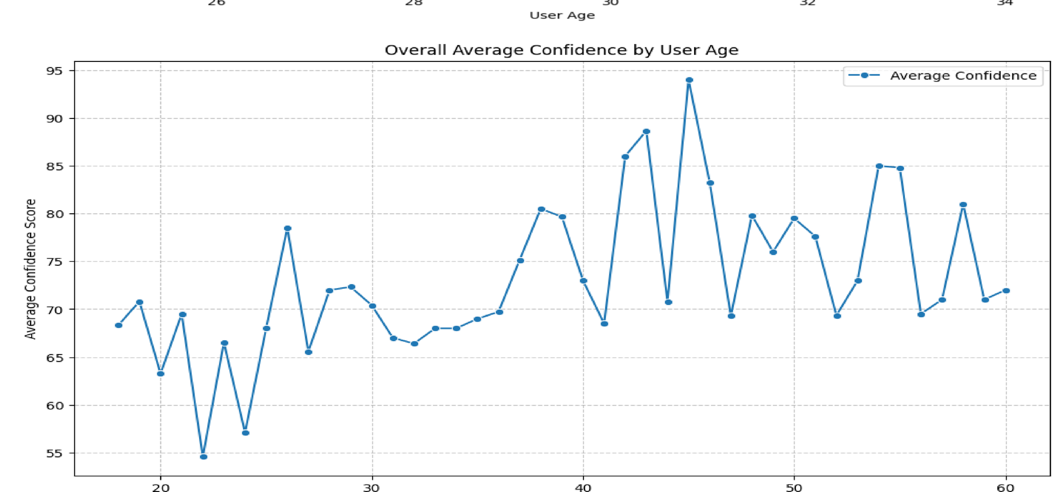
Fig 2. Confidence detection interface showing live analysis, questionnaire results, and confidence trend over time.

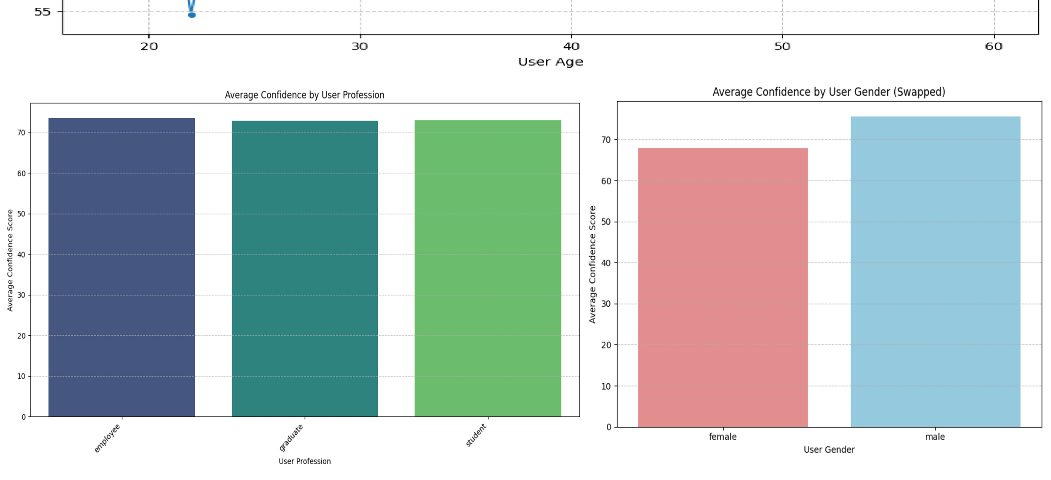
Results

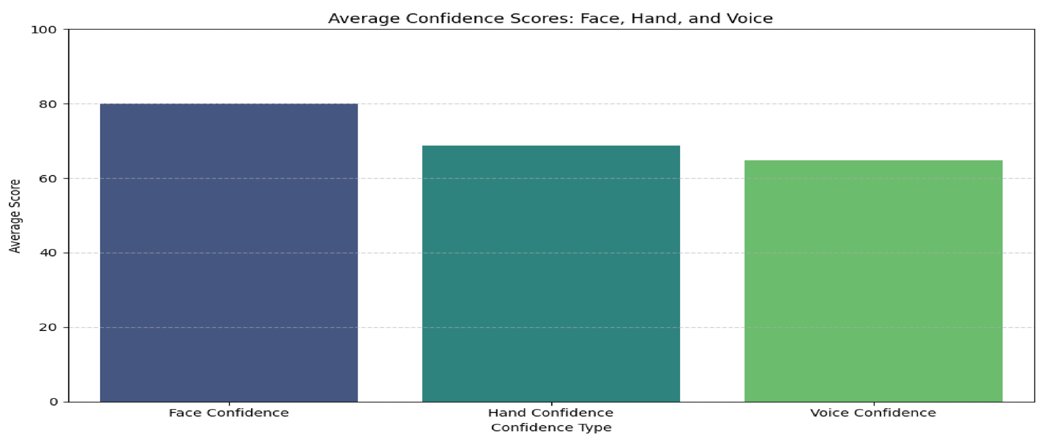
The results of the system demonstrate that facial expressions, hand gestures, and vocal characteristics all play significant roles in determining real-time confidence levels. Facial expression analysis showed that a smile with a lip aspect ratio greater than 1.5 increased confidence, while excessive blinking above 15 blinks per minute and head movement beyond ±10° noticeably reduced it. Consistent lip movement was associated with higher engagement, and a steady gaze strongly correlated with high confidence. Hand gesture analysis revealed that moderate gesture speeds between 0.2 and 0.5 m/s reflected calm and confident behavior, whereas rapid or erratic movements indicated low confidence. Vocal analysis further supported these findings, as stable pitch and balanced loudness were linked to high confidence, while high jitter, pitch fluctuation, and reduced vocal intensity suggested nervousness or uncertainty. When categorized using the overall scoring system, participants were grouped into high (0.9–1.2), medium (0.6–0.8), and low confidence levels (0.4–0.5). In terms of modality-based averages, facial cues showed the highest stability at 80.13%, followed by hand gestures at 68.81% and vocal features at 64.87%, reflecting the varying strengths of each modality in confidence detection.











Conclusion

The model that has been proposed is a very accurate way of detecting human confidence in real-time; it incorporates facial expressions, hand gestures, and vocal characteristics into one scoring system. The detection of behavioral cues like gaze stability, blink rate, gesture smoothness, and vocal clarity brings out the model's ability to give immediate feedback that is meaningful and can improve the quality of communication in virtual settings. It already has a strong impact on online interviews, public speaking practice and remote learning where sometimes user confidence is the main concern. The model is already quite accurate in the controlled settings but it can be improved further by using larger and more diverse datasets that would take into consideration the different types of user behaviors and cultural differences. In the end, the system sets up a good way for multimodal confidence assessment and also paves the way for future innovations in real-time human-computer interaction.

Future Work

- **Multi-face detection:** Enable the system to track and analyze multiple users simultaneously for group meetings, classrooms, and collaborative settings.
- **Advanced hand-gesture recognition:** Incorporate more detailed gesture classification to capture complex expressive movements often used during presentations or interviews.
- **Eyebrow movement analysis:** Add detection of subtle eyebrow dynamics to improve recognition of emotional cues such as doubt, emphasis, or surprise.
- **Enhanced visual feedback:** Introduce real-time overlays, including facial landmarks, gesture paths, and confidence indicators, to make feedback more intuitive for users.
- **Improved model accuracy:** Train the system with larger and more diverse datasets to handle variations in culture, lighting, and individual behavior.

References

1. M. Archana, C. Nitish, S. Harikumar, Real time face detection and optimal face mapping for online classes, in: Journal of Physics: Conference Series, Vol. 2161, IOP Publishing, 2022, p.012063.

2. C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, M. Shah, Deep learningbased human pose estimation: A survey, ACM Computing Surveys 56 (1) (2023) 1–37.

3. H.-J. Wang, T. Jin, Y.-T. Zhang, Application of iafsa-bp neural network in face orientation recognition, in: Journal of Physics: Conference Series, Vol. 1651, IOP Publishing, 2020, p. 012137.

4. L. Gao, Y. Xu, Face orientation recognition based on multiple facial feature triangles, in: 2012 International Conference on Control Engineering and Communication Technology, 2012, p. 175.

5. W. L. Passos, I. M. Quintanilha, G. M. Araujo, Real-time deep-learning-based system for facial recognition, Simp´ osio Brasileiro de Telecomunicac¸˜ oes e Processamento de Sinais (SBrT) 37 (2018) 895–899.

6. C. Espinosa Sandoval, Multiple face detection and recognition system design applying deep learning in web browsers, University of Arkansas ScholarWorks (2019). csceuht/74/

7. M. Rodriguez, C. Johnson, Anti-cheating mechanisms in online learning: A review of current techniques, Journal of E-Learning and Higher Education 2021 (2021) 1–10.

Acknowledgement

It was the guidance, feedback, and cooperation of the people involved that made possible the development of this model thus, we express our gratitude to all of them. First, we would like to thank the individuals who participated in testing and the team's support during data collection and evaluation of the system. Their input was crucial in maintaining fairness, accuracy, and integrity throughout the entire process. Furthermore, we would like to acknowledge the technical assistance and constructive suggestions which helped to improve the system's performance. Lastly, a big thank you to all who provided time, resources, and support to make this research happen.

Department of Computer Science, Faculty of Science and Technology.
American International University-Bangladesh (AIUB)