Multimodal AI-Driven Telehealth Platform For Emotion-Aware Diagnosis And Personalized Specialist Recommendation

Mrs. N. Savitha[1], M.Bharatha selvan[2], R.Surya[3], P.Ayus[4], S.Gnanashivan[5]
[1]Assist prof, Dept of Computer Science& Engineering,
[2, 3, 4, 5] Dept of Computer Science& Engineering,
[1, 2, 3, 4, 5] AVS Engineering College, Salem, India

Abstract- *Telehealth systems have significantly improved access to medical consultation, especially for patients in remote areas. However, existing systems often lack the ability to interpret patient emotions and provide personalized doctor recommendations. This paper proposes an advanced AI-based telehealth platform that enhances online medical consultation by integrating multiple intelligent techniques. The system employs deep learning-based facial expression analysis to detect patient emotions during video consultations. In addition, a speech recognition module converts patient speech into text, which is further processed using natural language processing (NLP) techniques to extract symptoms and identify possible medical conditions. Based on both emotional and clinical insights, an intelligent recommendation model suggests the most suitable doctor for the patient. The platform also supports secure video consultations, digital prescription generation, and integrated pharmacy delivery services. The proposed system aims to improve diagnostic accuracy, enhance patient-doctor interaction, and provide a more personalized and efficient healthcare experience.*

Keywords: Telehealth, Emotion Detection, Medical Text Analysis, Doctor Recommendation, Artificial Intelligence, Telemedicine

I. INTRODUCTION

Telehealth has emerged as a powerful solution for delivering healthcare services remotely using modern communication technologies. It enables patients to connect with healthcare professionals through video consultations, mobile applications, and remote monitoring systems, reducing the need for physical hospital visits. With the rapid growth of internet accessibility and smart devices, telehealth is becoming an essential part of the healthcare ecosystem. It not only improves accessibility for patients in rural and remote areas but also enhances the efficiency of healthcare delivery by saving time and reducing costs.

Despite these advancements, existing telehealth systems still face several limitations. Most platforms primarily rely on basic video or audio communication, which makes it difficult for doctors to fully understand a patient's emotional and psychological condition during consultations. Important factors such as facial expressions, speech patterns, and behavioral cues are often ignored, leading to incomplete diagnosis. In addition, patients are usually required to manually search for doctors, which may result in delays and mismatches in specialist selection.

Another major limitation of current systems is the lack of integration of advanced artificial intelligence techniques capable of analyzing multiple types of data simultaneously. Most existing solutions do not combine facial emotion recognition, speech analysis, and medical text understanding into a unified system. Furthermore, there is no complete pipeline that connects consultation, diagnosis, prescription, and medicine delivery in a seamless manner. As a result, patients still need to visit pharmacies separately, reducing the overall convenience of telehealth services.

To address these challenges, this paper proposes a Multimodal AI-Driven Telehealth Platform that integrates advanced machine learning and deep learning techniques to improve remote healthcare services. The system analyzes patient emotions using the Multimodal Emotion Transformer (MET), processes speech input through the Conformer model, and extracts clinical information using Med-BERT. Based on the combined analysis, the platform recommends the most suitable medical specialist using Neural Collaborative Filtering (NCF).

The main contribution of this work is the development of an intelligent and integrated telehealth system that enhances diagnosis accuracy and personalization. The proposed platform provides emotion-aware consultation, automated specialist recommendation, and seamless prescription-to-pharmacy integration for medicine delivery. By combining emotional and clinical insights, the system

improves the quality of healthcare services and offers a more efficient, accessible, and user-centered telehealth experience.

II. LITERATURE SURVEY

Telehealth has emerged as a powerful solution for delivering healthcare services remotely using modern communication technologies. It enables patients to connect with healthcare professionals through video consultations, mobile applications, and remote monitoring systems, reducing the need for physical hospital visits. With the rapid growth of internet accessibility and smart devices, telehealth is becoming an essential part of the healthcare ecosystem. It not only improves accessibility for patients in rural and remote areas but also enhances the efficiency of healthcare delivery by saving time and reducing costs.

Despite these advancements, existing telehealth systems still face several limitations. Most platforms primarily rely on basic video or audio communication, which makes it difficult for doctors to fully understand a patient's emotional and psychological condition during consultations. Important factors such as facial expressions, speech patterns, and behavioral cues are often ignored, leading to incomplete diagnosis. In addition, patients are usually required to manually search for doctors, which may result in delays and mismatches in specialist selection.

Another major limitation of current systems is the lack of integration of advanced artificial intelligence techniques capable of analyzing multiple types of data simultaneously. Most existing solutions do not combine facial emotion recognition, speech analysis, and medical text understanding into a unified system. Furthermore, there is no complete pipeline that connects consultation, diagnosis, prescription, and medicine delivery in a seamless manner. As a result, patients still need to visit pharmacies separately, reducing the overall convenience of telehealth services.

To address these challenges, this paper proposes a Multimodal AI-Driven Telehealth Platform that integrates advanced machine learning and deep learning techniques to improve remote healthcare services. The system analyzes patient emotions using the Multimodal Emotion Transformer (MET), processes speech input through the Conformer model, and extracts clinical information using Med-BERT. Based on the combined analysis, the platform recommends the most suitable medical specialist using Neural Collaborative Filtering (NCF).

The main contribution of this work is the development of an intelligent and integrated telehealth system that enhances diagnosis accuracy and personalization. The proposed platform provides emotion-aware consultation, automated specialist recommendation, and seamless prescription-to-pharmacy integration for medicine delivery. By combining emotional and clinical insights, the system improves the quality of healthcare services and offers a more efficient, accessible, and user-centered telehealth experience.

.

III. METHODOLOGY

The proposed system is a multimodal AI-driven telehealth platform designed to enhance remote healthcare services by integrating emotion detection, speech processing, medical text analysis, and intelligent doctor recommendation. The system processes multiple forms of patient data, including facial expressions, speech input, and textual medical information, to provide accurate and personalized healthcare support. The overall workflow begins with patient interaction and ends with prescription generation and medicine delivery.

Emotion Detection using MET

To understand the emotional state of patients during consultation, the system utilizes a Multimodal Emotion Transformer (MET). This model analyzes facial expressions captured through the patient's camera and identifies emotional cues such as stress, anxiety, or discomfort. The MET model combines visual features with contextual information to improve the accuracy of emotion recognition. By detecting emotional states, the system assists doctors in better understanding the patient's condition beyond verbal communication, leading to more effective diagnosis.

Speech Processing using Conformer

The system incorporates a Conformer-based model to process and analyze patient speech input. This model efficiently captures both local and global dependencies in audio signals by combining convolutional neural networks with transformer architectures. The patient's spoken symptoms are converted into text using automatic speech recognition (ASR) techniques. This enables the system to accurately interpret patient concerns even in cases where manual text input is not provided, improving usability and accessibility.

Medical Text Analysis using Med-BERT

For extracting meaningful clinical insights from textual data, the system employs Med-BERT, a domain-specific language model trained on medical datasets. It processes patient symptoms, medical history, and doctor notes

to identify key medical entities and conditions. Med-BERT enhances the understanding of unstructured medical text and helps in generating accurate clinical interpretations. This step plays a crucial role in bridging the gap between raw patient input and structured medical knowledge.

Doctor Recommendation using NCF

To provide personalized healthcare services, the system uses Neural Collaborative Filtering (NCF) for recommending suitable doctors. The recommendation model considers multiple factors such as patient symptoms, emotional state, medical history, and previous interactions. Based on these inputs, the system suggests the most appropriate specialist for the patient. This reduces the time required to find the right doctor and improves the overall efficiency of the consultation process.

Integrated Workflow

The proposed system follows a structured workflow that ensures seamless interaction between different components. Initially, the patient logs into the platform and provides input through video, speech, or text. The system simultaneously processes emotional, audio, and textual data using the respective AI models. The outputs from these models are combined to generate a comprehensive understanding of the patient's condition. Based on this analysis, the system recommends a suitable doctor, facilitates consultation, and generates a prescription. Finally, the prescription is forwarded to the pharmacy module for medicine preparation and delivery tracking.

IV. SYSTEM ARCHITECTURE

The system architecture of the proposed AI-driven telehealth platform is designed to provide a seamless and efficient healthcare workflow by integrating multiple modules, including patient interaction, AI processing, doctor recommendation, and pharmacy services. The architecture follows a structured pipeline in which data flows through different stages, enabling accurate analysis and decision-making.

Overview of Architecture

The overall architecture consists of three main layers: the user interaction layer, the AI processing layer, and the service layer. The user interaction layer includes patients, doctors, and pharmacy administrators who interact with the system through a web-based interface. The AI processing layer is responsible for analyzing patient data using advanced machine learning models, while the service layer handles consultation, prescription management, and medicine delivery.

Patient Interaction Module

The process begins with the patient logging into the system and providing input through multiple modes such as video, speech, and text. The video input is used to capture facial expressions, while speech input allows patients to describe their symptoms naturally. Additionally, patients can enter textual information regarding their medical history or current condition. This multimodal input ensures that the system gathers comprehensive data for analysis.

AI Processing Module

Once the patient data is collected, it is forwarded to the AI processing module, which consists of multiple specialized models. The facial data is processed using the Multimodal Emotion Transformer (MET) to detect emotional states. The speech input is analyzed using the Conformer model for accurate speech-to-text conversion. The extracted text, along with manually entered data, is processed using Med-BERT to derive meaningful clinical insights. These models work together to generate a detailed understanding of the patient's condition.

Recommendation and Consultation Module

Based on the processed data, the system uses Neural Collaborative Filtering (NCF) to recommend the most suitable doctor for the patient. The recommendation is personalized by considering symptoms, emotional condition, and historical data. Once a doctor is selected, the system enables real-time consultation through chat or video communication. This module ensures that patients receive timely and appropriate medical attention.

Prescription and Pharmacy Module

After the consultation, the doctor provides a digital prescription through the system. This prescription is automatically forwarded to the pharmacy module, where the required medicines are prepared. The system also supports order management, billing, and delivery tracking, ensuring that patients receive their medications without the need to visit a physical pharmacy.

Workflow Summary

The complete workflow of the system can be summarized as follows: the patient provides multimodal input,

which is analyzed using AI models to understand both emotional and clinical aspects. Based on this analysis, the system recommends a suitable doctor and facilitates consultation. Finally, the prescription is generated and processed by the pharmacy module for medicine delivery. This integrated architecture ensures a smooth and efficient healthcare experience.

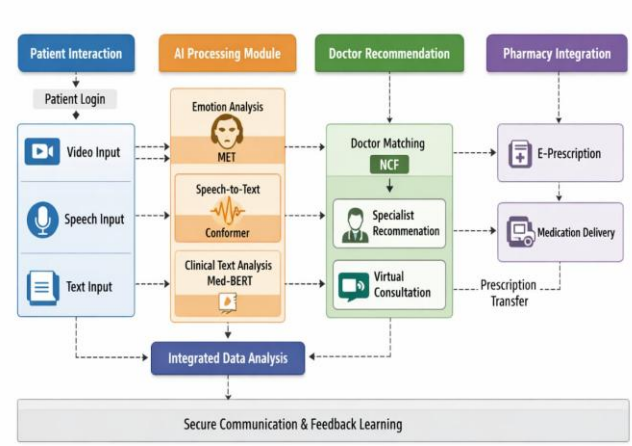

Fig 1 . System Architecture of the Proposed AI-Driven Telehealth Platform

V. RESULTS

The proposed AI-driven telehealth platform was tested under various scenarios to evaluate its performance, accuracy, and usability. The system integrates multiple modules, including emotion detection, speech processing, medical text analysis, and doctor recommendation, to provide a comprehensive healthcare solution. The results demonstrate that the system performs efficiently in processing multimodal data and delivering accurate outputs in real-time.

The emotion detection module, based on the Multimodal Emotion Transformer (MET), was able to identify patient emotions such as stress, anxiety, and neutral states with satisfactory accuracy. This feature significantly improves the understanding of patient conditions during remote consultations. Similarly, the Conformer-based speech processing module effectively converted patient speech into text with high reliability, enabling seamless communication even without manual input.

The Med-BERT model used for medical text analysis successfully extracted relevant clinical information from patient inputs, including symptoms and medical history. This improved the system's ability to interpret unstructured data and assist in clinical decision-making. Furthermore, the Neural Collaborative Filtering (NCF) model provided accurate doctor recommendations by analyzing patient data and preferences, reducing the time required to find suitable specialists.

The integration of these modules resulted in an efficient end-to-end workflow, where patients could complete the entire process from consultation to prescription and medicine delivery within a single platform. The system also demonstrated good responsiveness and user-friendly interaction, making it suitable for real-world applications. Compared to traditional telehealth systems, the proposed solution offers enhanced personalization, improved diagnostic support, and better accessibility.

Overall, the results indicate that the proposed system is capable of delivering intelligent and reliable healthcare services by leveraging multimodal AI techniques. The discussion highlights that integrating emotion, speech, and text analysis provides a more holistic understanding of patient conditions, thereby improving the quality and effectiveness of remote healthcare services.

VI. CONCLUSION

This paper presented a multimodal AI-driven telehealth platform designed to enhance the efficiency and quality of remote healthcare services. The proposed system integrates advanced machine learning models, including the Multimodal Emotion Transformer (MET), Conformer, Med-BERT, and Neural Collaborative Filtering (NCF), to analyze patient data from multiple sources such as facial expressions, speech, and medical text. By combining these technologies, the system provides a more comprehensive understanding of patient conditions compared to traditional telehealth solutions.

The implementation of emotion-aware analysis enables healthcare professionals to better interpret the psychological state of patients during consultations, leading to improved diagnosis and decision-making. In addition, the intelligent doctor recommendation system ensures that patients are connected with the most suitable specialists in a timely manner. The integration of consultation, prescription generation, and pharmacy services into a single platform further enhances user convenience and reduces the overall healthcare process time.

The results demonstrate that the proposed system is capable of delivering accurate, efficient, and personalized healthcare services. By leveraging multimodal AI techniques, the platform addresses key limitations of existing systems and provides a more reliable and user-centered telehealth experience. Overall, the proposed solution contributes to the

advancement of digital healthcare by improving accessibility, automation, and quality of care.

VII. FUTUREWORK

Although the proposed AI-driven telehealth platform demonstrates effective performance, there are several opportunities for further improvement and expansion. Future work can focus on enhancing the accuracy of emotion detection by incorporating more advanced deep learning models and larger real-time datasets. Integrating wearable devices such as smartwatches and health monitoring sensors can enable continuous tracking of vital parameters like heart rate, blood pressure, and oxygen levels, providing more comprehensive patient data.

In addition, the system can be extended to support mobile-based applications for improved accessibility and user convenience. The inclusion of multilingual support can make the platform more user-friendly for patients from diverse linguistic backgrounds. Further improvements can also be made in the recommendation system by incorporating real-time feedback and adaptive learning techniques to provide more precise and personalized doctor suggestions.

Another potential enhancement is the integration of cloud-based infrastructure and blockchain technology to ensure secure storage and sharing of medical records. This can improve data privacy, transparency, and reliability. Moreover, incorporating advanced predictive analytics can help in early disease detection and preventive healthcare.

Overall, future developments aim to make the system more scalable, secure, and intelligent, thereby expanding its applicability in real-world healthcare environments and improving the quality of telehealth services.

VIII. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to their project guide for providing valuable guidance, continuous support, and constructive suggestions throughout the development of this work. The successful completion of this project would not have been possible without their encouragement and expert advice.

The authors also extend theirthanks to the faculty members and institution for offering the necessary resources and a supportive environment to carry out this research. Their contributions have played a vital role in the successful implementation of the proposed system.

Finally, the authors would like to thank their friends and family for their constant motivation and support during the course of this project.

Figure and Experimental Visualization

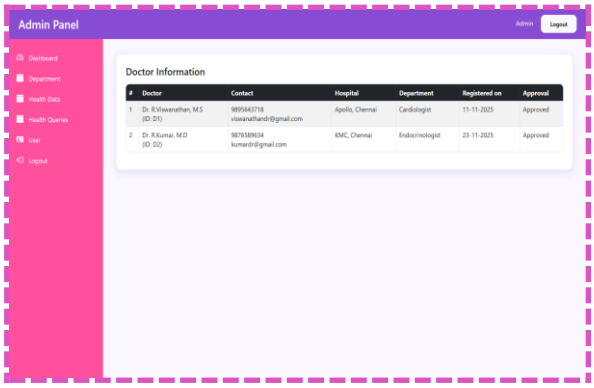

Fig 2.1 Admin panel

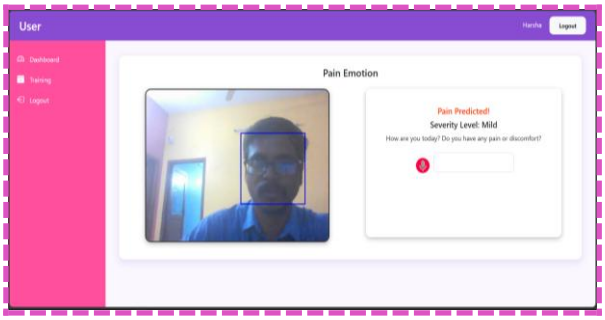

Fig2.2 patient panel

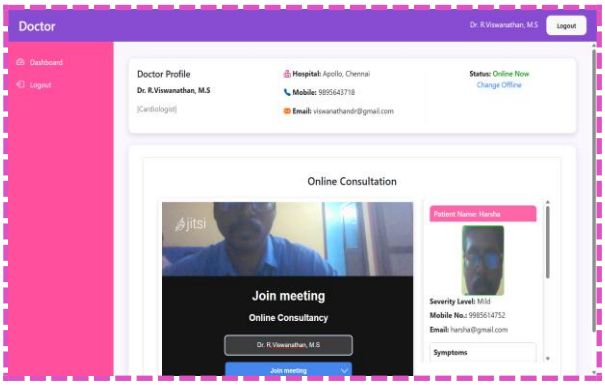

Fig 2.3 Doctor & patient Meeting

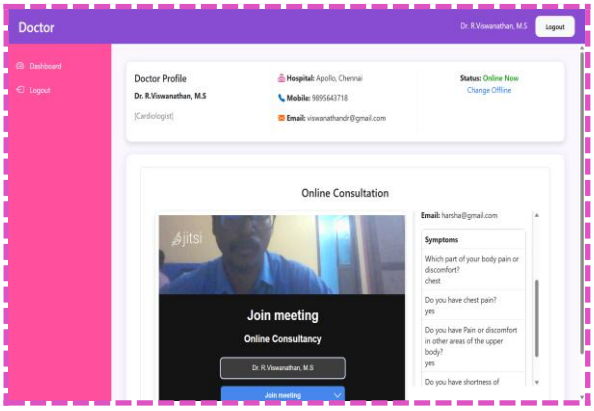

Fig 2.4 Doctor Prescribed Medication

REFERENCES

[1] X. Kou, "Personalized doctor recommendation method based on patient online consultation text," 2021.

[2] J. Guo, Y. Dong, X. Liu, and S. Lu, "Facial expression recognition improved by attention mechanism and involution operator," *Computer Engineering and Applications*, vol. 59, no. 23, pp. 95–103, 2023.

[3] Y.-F. Huang, P. Liu, Q. Pan, and J.-S. Lin, "A doctor recommendation algorithm based on doctor performances and patient preferences," 2012.

[4] C. Shi, C. Tan, and L. Wang, "A facial expression recognition method based on a multibranch cross-connection convolutional neural network," *IEEE Access*, vol. 9, pp. 39255–39274, 2021.

[5] C. Yi, L. Chusong, and T. Minglu, "Online doctor recommendation method for doctor–patient interaction data and doctor attributes," 2023.

[6] R. Capozzo *et al.*, "Telemedicine is a useful tool to deliver care to patients with Amyotrophic Lateral Sclerosis during COVID-19 pandemic: Results from Southern Italy," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 21, no. 7–8, pp. 542–548, Nov. 2020.

[7] H. Jiang and W. Xu, "How to find your appropriate doctor: An integrated recommendation framework in big data context," 2014.

[8] S. Williams *et al.*, "The discerning eye of computer vision: Can it measure Parkinson's finger tap bradykinesia," *Journal of the Neurological Sciences*, vol. 416, 2020.

[9] C. Ju and S. Zhang, "Doctor recommendation model for pre-diagnosis online in China: Integrating ontology characteristics and disease text mining," 2021.

[10] C. Duncan *et al.*, "Video consultations in ordinary and extraordinary times," *Practical Neurology*, vol. 20, no. 5, pp. 396–403, Oct. 2020.