IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

AI-Driven Healthcare System For Doctor Recommendation And Video Consultation Based On Facial Expression And Speech Analysis

[1]Saranya V, [2]Arul Maria Agnes I, [3]Ginu K, [4]Gowtham S
[1]Asst. Professor, [2]Student, [3]Student, [4]Student
[1]Computer Science And Engineering,
[1]Adhiyamaan College Of Engineering, Hosur, India

Abstract : Ai-driven healthcare system for doctor recommendation and video consultation based on facial expression and speech analysis employs advanced technologies to enhance patient care, focusing on emotional well-being. The system combines Temporal Convolutional Neural Networks (TCNN) for facial expression analysis, Convolutional Neural Networks (CNN) for speech recognition, Natural Language Processing (NLP) for semantic understanding, and content-based filtering to recommend healthcare professionals. The core of the system lies in its ability to provide video consulting services, allowing real-time monitoring of patients' emotional and mental states during remote consultations.

I. INTRODUCTION

Remote healthcare services, or telehealth, have transformed the medical industry by enabling remote consultations through digital technologies. However, many existing systems struggle to accurately interpret patients' emotional states, which can impact diagnosis and treatment. Current telehealth platforms often focus on physical symptoms while neglecting emotional well-being, leading to incomplete assessments. Additionally, conventional doctor recommendation systems may lack personalization, making it difficult for patients to connect with the most suitable healthcare professional. To address these limitations, an AI-driven solution is needed to enhance remote healthcare by integrating advanced emotional analysis and personalized consultations. The proposed system leverages Temporal Convolutional Networks (TCN) for real-time facial expression analysis, Convolutional Neural Networks (CNN) for speech recognition, and Natural Language Processing (NLP) for understanding patient concerns through text analysis. It also incorporates content-based filtering to provide personalized doctor recommendations by analyzing a patient's emotional state, speech patterns, and medical needs. By combining these technologies, the system ensures a more holistic and accurate approach to patient care during video consultations, enhancing both emotional and physical health assessments.

With real-time emotional monitoring, healthcare providers gain deeper insights into a patient's mental state, enabling more empathetic and effective consultations. The system's multi-modal analysis integrates facial expressions, speech, and text to offer a comprehensive health overview, while personalized doctor recommendations ensure patients connect with the right specialists. Additionally, the user-friendly interface makes remote healthcare accessible and efficient. By integrating AI-driven capabilities, this system redefines telehealth, making it more personalized, accurate, and emotionally aware.

1.1 Data Collection

This study, ensuring the availability of diverse and high-quality datasets to train the AI-driven healthcare system effectively. The system relies on multiple sources, including publicly available datasets, real-world patient consultations, and simulated healthcare interactions. Facial expression data is sourced from well-established datasets such as FER-2013, CK+, and AffectNet, which contain labeled images representing various emotions like happiness, sadness, anger, and surprise. These datasets are instrumental in training the Temporal Convolutional Network (TCN) for real-time emotion recognition. Additionally, speech data is obtained from repositories such as RAVDESS, LibriSpeech, and Emo-DB, which provide audio samples annotated with emotional tones. Speech features are extracted using Mel-Frequency Cepstral Coefficients (MFCCs) and wavelet transformations, which help the system classify verbal cues and infer patient emotions accurately. Text data is collected from medical dialogue datasets like MIMIC-III and PubMed, allowing Natural Language Processing (NLP) models to analyze patient queries and extract meaningful insights from consultations. Furthermore, real-time patient-doctor interaction data is gathered from telehealth platforms with patient consent, enabling the recommendation system to personalize doctor suggestions based on patient history, symptoms, and emotional states. Ethical considerations are strictly adhered to, with data anonymization techniques ensuring patient privacy and compliance with regulations such as HIPAA and GDPR. User consent is obtained before collecting any real-world data, reinforcing the study's commitment to ethical AI deployment.

1.2 Preprocessing and Analysis

To ensure optimal model performance, raw data undergoes extensive preprocessing, including cleaning, transformation, and feature extraction. Facial expression data is processed by resizing images to standardized dimensions, normalizing pixel values, and applying noise reduction techniques such as Gaussian filtering. Data augmentation techniques, including rotation, flipping, and brightness adjustments, are used to enhance model generalization and reduce overfitting. Speech data is preprocessed using wavelet-based noise filtering to remove background disturbances, while Mel-Frequency Cepstral Coefficients (MFCCs) are extracted to capture tone and pitch variations. Text data is cleaned through tokenization, lemmatization, and stopword removal, allowing the NLP model to focus on meaningful medical terms. Sentiment analysis and Named Entity Recognition (NER) techniques are applied to extract medical symptoms and emotional cues from patient interactions. For the doctor recommendation system, patient profiles and medical history are structured to enable efficient feature matching. Cosine similarity and TF-IDF techniques are employed to measure patient-doctor compatibility, ensuring accurate and context-aware recommendations.

1.3 Modeling Approach

The system integrates multiple deep learning models to enable efficient emotion recognition, speech processing, and doctor recommendations. Facial expression recognition is performed using a Temporal Convolutional Network (TCN), which captures sequential patterns in facial expressions over time, enabling real-time emotional monitoring during consultations. The speech recognition model employs a CNN-RNN hybrid, where CNN layers extract phonetic features from audio while RNN (LSTM/GRU) layers process temporal dependencies in speech. This ensures accurate speech-to-text conversion and tone analysis for sentiment detection. The NLP-based sentiment and symptom analysis module leverages a BERT-based NLP model to extract medical insights from patient interactions, identifying symptoms and emotional states through Named Entity Recognition and contextual analysis. The doctor recommendation system adopts a hybrid approach, combining content-based filtering and collaborative filtering to match patients with suitable specialists. Content-based filtering maps symptoms to medical expertise, while collaborative filtering incorporates patient feedback to refine recommendations over time. By integrating these AI-driven techniques, the system enhances diagnostic accuracy and personalization in remote healthcare.

1.4 Model Evaluation

Each model is rigorously evaluated using standard performance metrics to ensure accuracy, reliability, and real-world applicability. The facial expression recognition model is assessed using accuracy, F1-score, and confusion matrices to measure classification effectiveness. Cross-validation is performed to ensure generalization across different facial expressions. The speech recognition model is evaluated using Word Error Rate (WER), Signal-to-Noise Ratio (SNR), and phoneme accuracy to ensure high-quality speech-to-text conversion. For NLP-based sentiment and symptom analysis, BLEU scores, precision-recall metrics, and sentiment correlation tests measure the effectiveness of emotion detection and medical text understanding.

The doctor recommendation system is validated using Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) to assess ranking quality and recommendation accuracy. Additionally, user feedback scores collected from real-world telehealth sessions provide insights into system usability and effectiveness, allowing continuous model refinement.

II. RESEARCH METHODOLOGY

The study follows a structured research methodology that ensures the systematic development, evaluation, and optimization of the AI-driven healthcare system. Initially, a literature review is conducted to identify gaps in existing telehealth solutions, particularly in emotion detection and personalized recommendations. The data collection phase involves acquiring diverse datasets from facial expression databases, speech recordings, medical text sources, and real-world telehealth interactions. Preprocessing and feature engineering techniques are applied to enhance data quality before model training. Deep learning models are developed using supervised learning and transfer learning approaches, enabling efficient training and real-world adaptation. The evaluation phase involves rigorous testing using statistical performance metrics and real-world validation in telehealth environments. Finally, the system is deployed as a functional web-based telehealth application, ensuring seamless interaction between patients and healthcare providers. An iterative refinement process based on user feedback and continuous learning ensures sustained improvements in system accuracy and usability.

2.1 Theoretical framework

The research is grounded in theories from artificial intelligence, healthcare informatics, and human-computer interaction. Deep learning principles, including convolutional and recurrent neural networks, underpin the emotion recognition, speech processing, and NLP models. Reinforcement learning strategies are integrated into the doctor recommendation system to refine suggestions based on patient feedback. Cognitive psychology theories on emotional recognition provide insights into facial expression and sentiment analysis, enhancing the AI's ability to understand human emotions. Healthcare informatics frameworks guide the development of telemedicine solutions, ensuring compliance with medical best practices and ethical considerations. Human-computer interaction (HCI) principles inform the design of the telehealth platform's user interface, optimizing usability and accessibility for patients and healthcare providers. By integrating these theoretical foundations, the AI-driven system fosters a patient-centric, emotionally intelligent telehealth experience.

III. MODEL DEVELOPMENT

The development of the AI-driven healthcare system involves implementing multiple deep learning models for facial expression recognition, speech analysis, Natural Language Processing (NLP)-based sentiment detection, and personalized doctor recommendations. Each of these models is designed to work in real-time to ensure seamless integration within telehealth consultations. The system follows a modular approach, where different AI components collaborate to enhance patient assessment and deliver an emotionally intelligent telehealth experience. By leveraging advanced deep learning techniques, the system aims to bridge the gap between conventional telehealth and a more personalized, AI-assisted consultation process.

3.1 Facial Expression Recognition Model

The facial expression recognition model is built using **Temporal Convolutional Networks (TCN)** to analyze a patient's emotional state during video consultations. The model is trained using well-established datasets such as **FER-2013, CK+, and AffectNet**, which provide a diverse set of labeled facial images representing different emotional states, including happiness, sadness, anger, and fear. The architecture consists of convolutional layers for feature extraction, followed by temporal layers to capture sequential variations in facial expressions. The fully connected layers classify the detected emotions into predefined categories, ensuring real-time emotional monitoring. Activation functions like **ReLU** are used for feature extraction, while **Softmax** is applied in the final layer for multi-class emotion classification. The loss function employed is **cross-entropy**, optimizing the accuracy of emotion detection. To enhance model robustness, **data augmentation** techniques such as image rotation, flipping, and brightness adjustments are used to prevent overfitting. Additionally, **face alignment and Region Proposal Networks (RPNs)** are incorporated to focus on key facial landmarks, ensuring that the system accurately detects and interprets emotional cues in real-world scenarios.

3.2 Speech Recognition Model

The speech recognition module enables speech-to-text conversion while also analyzing the emotional tone present in a patient's voice. This is achieved through a CNN-RNN hybrid model, where Convolutional Neural Networks (CNNs) are used for feature extraction and Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), capture the sequential dependencies of speech signals. The preprocessing pipeline begins with wavelet filtering, which removes background noise and enhances speech clarity. Following this, Mel-Frequency Cepstral Coefficients (MFCCs) are extracted to analyze tone, pitch, and phonetic features. The CNN layers process the extracted spectrograms, while the RNN layers help in understanding the temporal nature of speech. The final softmax layer converts the output into textual representation, making the system capable of understanding spoken language. The model is trained on datasets such as LibriSpeech, RAVDESS, and Emo-DB, which provide emotionally annotated speech samples, enabling the system to detect signs of stress, anxiety, and emotional distress in a patient's voice. By incorporating sentiment analysis, the system determines whether the patient is speaking in a calm, distressed, or anxious tone, adding another layer of emotional intelligence to the telehealth consultation process.

3.3 NLP-Based Sentiment and Symptom Analysis Model

The NLP-based sentiment and symptom analysis module is designed to interpret patient concerns expressed in textual or spoken form. The system utilizes a Bidirectional Encoder Representations from Transformers (BERT)-based NLP model, which allows it to understand complex linguistic structures and extract relevant medical insights from patient conversations. The workflow begins with tokenization, where patient speech or text input is broken down into meaningful words and phrases. Named Entity Recognition (NER) is then applied to extract medical terms, symptoms, and conditions, helping the system detect early signs of health issues. The sentiment analysis component classifies the patient's emotional state, identifying whether they exhibit stress, urgency, or calmness based on their textual input. Through semantic similarity techniques, the system maps the extracted symptoms to relevant medical conditions, ensuring a context-aware understanding of the patient's situation. The model is trained on datasets such as MIMIC-III and PubMed medical text corpora, ensuring that it can accurately analyze patient symptoms and medical queries. This NLP integration allows the system to assist healthcare professionals in diagnosing conditions more effectively by providing emotionally aware textual analysis.

3.4 Doctor Recommendation System Model

The doctor recommendation system leverages hybrid filtering to suggest the most suitable medical professionals based on the patient's symptoms, emotional state, and past consultation history. This recommendation engine combines Content-Based Filtering (CBF) and Collaborative Filtering (CF) to provide accurate and personalized doctor suggestions. Content-Based Filtering works by analyzing patient symptoms and matching them with the expertise of available doctors using TF-IDF vectorization and cosine similarity techniques. This ensures that patients are recommended specialists whose expertise aligns with their medical concerns. Collaborative Filtering, on the other hand, enhances recommendations by incorporating patient feedback, consultation history, and ratings to improve system accuracy over time. By merging explicit patient preferences with AI-driven symptom matching, the hybrid recommendation model continuously refines doctor matching through real-time feedback and updates. This approach ensures that patients receive highly relevant, data-driven doctor recommendations, improving the efficiency of telehealth services.

3.5 Integration and Deployment

Once trained, the models are deployed within an AI Healthcare Consultant Web Application, which serves as the central hub for real-time telehealth services. The platform is developed using Python, Flask, MySQL, TensorFlow, and Keras, ensuring a scalable and efficient backend infrastructure. The frontend is designed using HTML, CSS, and Bootstrap, providing a user-friendly interface for both patients and healthcare professionals. The backend is managed using Flask, which handles AI model interactions and user requests. A MySQL database stores user profiles, medical history, and doctor recommendations, ensuring secure data management. Real-time model execution is enabled through TensorFlow Serving, allowing for fast and efficient AI inference during live consultations. The integration of these AI models within a telehealth platform enables emotionally aware doctor-patient interactions, allowing healthcare professionals to assess not only physical symptoms but also the emotional well-being of patients. This approach significantly enhances telehealth consultations, making them more human-centric, personalized, and responsive to patient emotions.

IV. RESULTS AND DISCUSSION

The descriptive statistics provide an overview of the key variables in the study, including emotion recognition accuracy, speech analysis performance, sentiment classification, and doctor recommendation accuracy. These results summarize the mean, standard deviation, minimum, and maximum values of the AI models used in the telehealth system. The analysis helps in understanding the performance trends and areas requiring further improvement.

4.1 Results of Descriptive Statics of Study Variables

Table 4.1: Descriptive Statics

Parameters	Mean	Standard Deviation	Minimum	Maximum
Hapiness	94.2	1.5	91.0	96.5
Sadness	92.7	2.0	89.5	95.3
Anger	91.3	2.5	87.2	94.8
Fear	89.8	3.1	85.4	93.6
Word Error Rate (WER)	7.8	1.1	6.1	9.5
Sentiment Detection Accuracy	89.5	2.3	85.0	92.8
Neutral Speech Detection	93.2	1.8	90.4	96.5
Stress/Anxiety Detection	88.4	2.7	83.6	91.7
Named Entity Recognition (NER) Accuracy	91.4	2.2	88.0	94.7
Precision-Recall Score	89.0	2.5	85.2	92.5
Sentiment Classification (Stress vs. Calm)	90.1	2.0	86.7	93.4
User Satisfaction Rate	91.3	2.4	87.5	94.2
Consultation Match Improvement (Post-Feedback Refinement)	8.4	1.1	6.9	9.8

Table 4.1 presents the descriptive statistics for various parameters related to emotion recognition, speech analysis, sentiment detection, and user satisfaction. The table provides insights into the model's overall performance by summarizing the mean, standard deviation, minimum, and maximum values for each parameter. The emotion recognition metrics indicate that happiness (94.2%) and sadness (92.7%) are detected with high accuracy, while fear (89.8%) shows slightly lower performance. The standard deviation values suggest some variability in the model's predictions, with anger having the highest variation (2.5). The Word Error Rate (WER) of 7.8% indicates a relatively low level of transcription errors, ensuring the model's speech recognition capabilities remain reliable.

Sentiment detection and speech classification also demonstrate strong performance, with neutral speech detection at 93.2% and sentiment detection accuracy at 89.5%. However, stress and anxiety detection show slightly higher variability (standard deviation of 2.7), indicating possible misclassification in distinguishing stress levels. Additionally, Named Entity Recognition (NER) achieves a mean accuracy of 91.4%, ensuring the model effectively identifies key terms in user inputs. The precision-recall score of 89.0% and sentiment classification at 90.1% further confirm that the model balances relevance and false positive rates effectively.

User experience-related metrics show promising results, with a user satisfaction rate of 91.3%, reflecting high acceptance and reliability of the system. Moreover, consultation match improvement, which measures the refinement in recommendation accuracy after user feedback, stands at 8.4%, indicating the system's adaptability in improving doctor-patient matching over time. While most parameters exhibit high accuracy and reliability, slight variations in stress detection and consultation matching suggest potential areas for further refinement. These insights help in identifying key strengths and areas for future enhancements in the model's performance.

V. ACKNOWLEDGMENT

We sincerely express our gratitude to our institution for providing the necessary resources and guidance for this research. We extend our appreciation to our mentors and faculty members for their valuable support and insightful feedback, which greatly contributed to our work.

We also acknowledge the contributions of researchers whose work provided valuable insights. Lastly, we are grateful to our families and friends for their continuous support and encouragement throughout this journey.

REFERENCES

[1] J. Guo, Y. Dong, X. Liu and S. Lu, "Facial expression recognition improved by attention mechanism and involution operator", Comput. Eng. App.,IEEE Access, vol. 59, no. 23, pp. 95-103, 2023.

[2] 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.

[3] 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 Scler. Frontotemporal Degeneration, IEEE Access,vol. 21, no. 7-8, pp. 542-548, Nov. 2020

[4] S. Williams et al., "The discerning eye of computer vision: Can it measure Parkinson's finger tap bradykinesia", J. Neurological Sci., IEEE Access,vol. 416, 2020.

[5] C. Duncan et al., "Video consultations in ordinary and extraordinary times", Practical Neurol., IEEE Access,vol. 20, no. 5, pp. 396-403, Oct. 2020.