Developing an Interactive 3D Digital Human for Real-Time Emotional Support and Therapy

Dr. Abdul Khadeer, [1]Assosiate Professor, Department of Computer Science and Technology,

Deccan College of Engineering and Technology, Hyderabad, India

Suhana ,[2]PG Scholar, Department of Computer Science and Engineering,

Deccan College of Engineering and Technology, Hyderabad, India

ABSTRACT -- Mental health disorders such as stress, anxiety, depression, and loneliness have become increasingly prevalent due to rapid urbanization, academic pressure, demanding work environments, and changing lifestyles. Consequently, many individuals do not receive timely emotional assistance. Recent advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), Human-Computer Interaction (HCI), and Computer Graphics have created new opportunities for developing intelligent virtual assistants capable of providing accessible and immediate emotional support.

This project, "Developing an Interactive 3D Digital Human for Real-Time Emotional Support and Therapy," presents an AI-powered virtual therapist designed to offer empathetic and engaging conversations through both text and voice interaction. The proposed system integrates a realistic three-dimensional VRM-based avatar with conversational AI to simulate a human-like therapeutic experience. The avatar performs synchronized lip movements, facial expressions, eye blinking, breathing animations, and subtle body gestures, making interactions more natural and emotionally expressive. The proposed system demonstrates the effective integration of Artificial Intelligence, Natural Language Processing, speech technologies, and real-time 3D avatar animation into a unified mental wellness platform.

Index Terms—Artificial Intelligence, Emotional Support, Conversational AI, Natural Language Processing, Emotion Detection, Virtual Therapist, 3D Digital Human, Human-Computer Interaction, Speech Recognition, Speech Synthesis, VRM Avatar, 3D Avatar Animation, Facial Expression, Lip Synchronization, Real-Time Interaction, Mental Wellness.

1. INTRODUCTION

1.1 Background

Mental health and emotional well-being have become important concerns in modern society as individuals increasingly experience stress, anxiety, loneliness, academic pressure, demanding work environments, and other emotional challenges. Although professional counselling and psychological services provide valuable support, access to such services may be affected by factors such as cost, availability, geographical limitations, waiting periods, and social hesitation. These challenges create a need for accessible digital platforms that can provide immediate and continuous emotional assistance while encouraging users to communicate their feelings in a comfortable environment.

Conventional digital mental-health applications and conversational systems generally rely on text-based interaction to communicate with users. While text-based chatbots can provide immediate responses and basic emotional assistance, they often provide limited non-verbal communication and may not reproduce the natural characteristics of human interaction. The absence of facial expressions, speech, body movement, and visual representation can reduce the sense of presence and engagement during a conversation. These limitations create an opportunity to develop more interactive systems that combine conversational intelligence with multimodal communication and human-like visual representation.

The development of **Artificial Intelligence (AI)** and **Natural Language Processing (NLP)** has enabled computer systems to understand user input and generate contextually appropriate conversational responses. In addition, **Speech Recognition** and **Speech Synthesis** technologies allow users to communicate through voice rather than relying exclusively on typed messages. The integration of these technologies with **Human-Computer**

Interaction (HCI) and real-time computer graphics provides an opportunity to create digital agents capable of interacting with users through both verbal and visual communication.

This research proposes an **Interactive 3D Digital Human for Real-Time AI-Assisted Emotional Support**. The system integrates conversational AI with text and voice interaction to provide supportive responses to users. A **VRM-based three-dimensional avatar** is used to represent the virtual therapist and provide visual feedback during conversations. The avatar incorporates facial expressions, synchronized lip movements, eye blinking, breathing animations, and body gestures to create a more natural and engaging interaction experience. The system also performs conversational emotion estimation to identify emotional states from user input and uses the detected state to influence the avatar's response and expression.

The proposed platform further incorporates **dashboard monitoring and session management** to provide additional information about the interaction. The dashboard presents parameters such as the detected emotional state, estimated stress level, mood progression, session duration, and message count. Session information, including conversation history, emotional patterns, and completed activities, can be maintained for subsequent review. The system also provides automated **PDF-based session report generation,** allowing the collected information to be organized into a structured summary.

The proposed system is evaluated through functional and software testing to examine the operation of major components, including user interaction, text and voice communication, AI-generated responses, emotion detection, avatar animation, dashboard updates, session management, and report generation. Performance and usability considerations are also examined to determine whether the integrated platform operates reliably and provides an interactive user experience. The project documentation describes static, structural, and behavioral testing as the principal testing approaches used for the developed system.

Overall, the study integrates **conversational artificial intelligence, emotion-aware interaction, speech technologies, and real-time three-dimensional avatar animation** into a unified digital emotional-support platform. Rather than limiting interaction to conventional text-based communication, the proposed approach combines conversational responses with speech and expressive avatar behaviour while providing emotional visualization and session-level reporting. The objective is to develop an accessible and engaging AI-assisted platform that can support users through real-time emotional interaction while serving as a complementary technology rather than a replacement for professional mental-health services.

1.2 Research Gap

Existing AI-based mental-health applications have demonstrated the potential of conversational systems to provide accessible emotional support. However, many existing systems primarily depend on text-based interaction and provide limited speech communication, visual representation, non-verbal expression, and emotional interaction. Research on conversational agents has also identified challenges related to emotional intelligence, personalization, contextual interaction, and realistic non-verbal communication.

A further gap exists between **conversational AI and embodied interaction.** Conventional chatbots generally provide responses through text interfaces, whereas the proposed system combines conversational AI with a **three-dimensional VRM avatar** capable of facial expressions, lip synchronization, blinking, breathing, and body gestures. The system additionally incorporates emotion estimation, dashboard-based emotional monitoring, session management, and automated report generation.

Therefore, this research addresses the need for an **integrated AI-assisted emotional-support platform** that combines conversational interaction, speech technologies, emotion-aware processing, and an expressive 3D digital human within a single system.

1.3 Research Objectives

The study focuses on the following objectives:

1. **To develop an AI-assisted emotional-support system** capable of providing interactive conversational responses.

2. **To integrate text and voice interaction** using speech recognition and speech synthesis technologies.
3. **To implement conversational emotion detection** for identifying emotional states from user input.
4. **To develop a VRM-based 3D digital human** with facial expressions, lip synchronization, blinking, breathing, and body gestures.
5. **To provide real-time emotional and session monitoring** through a dashboard containing mood, stress, session, and interaction information.
6. **To implement session management and automated PDF reporting** for maintaining and summarizing interaction data.
7. **To evaluate the developed system** based on functional performance, response quality, usability, interaction reliability, and system stability.

1.4 Limitations of the Study

The proposed system has several limitations. First, the current emotion-detection component uses **keyword-based conversational emotion classification**, which may not capture complex emotional states, sarcasm, or subtle changes in user behaviour.

Second, the system is designed as an **AI-assisted emotional-support platform** and has not been clinically validated as a replacement for professional psychological treatment. The quality of generated responses may also depend on the conversational AI service and the information provided by the user.

The current implementation uses **browser-based Local Storage** for maintaining session information, which limits long-term and multi-device data management. The system also depends on browser speech capabilities and available computational resources for smooth voice and 3D-avatar interaction.

Finally, advanced multimodal emotion recognition using facial expressions, vocal characteristics, or physiological signals is not included in the current implementation.

1.5 Rationale of the Study

The increasing need for accessible emotional support motivates the development of intelligent systems that can provide immediate and engaging interaction. Although conversational AI can generate supportive responses, text-only interfaces may provide limited emotional and non-verbal interaction.

The proposed study addresses this limitation by combining **conversational AI, speech interaction, emotion detection, and an expressive 3D avatar** within a unified platform. The addition of dashboard monitoring and automated session reporting further enables users to review emotional and conversational information collected during an interaction.

Overall, the study is motivated by the need for a **more interactive and human-like AI-assisted emotional-support system** that extends conventional chatbot interaction through voice communication, visual embodiment, emotional awareness, and session-level analysis.

2. LITERATURE REVIEW

The increasing adoption of artificial intelligence, conversational systems, and digital interaction technologies has created new opportunities for providing accessible emotional support through computer-based applications. AI-powered mental-health chatbots have been explored as alternatives for delivering preliminary psychological assistance, particularly because they can provide continuous interaction and allow users to communicate without the hesitation that may occur in conventional settings. However, many existing systems primarily rely on text-based communication, which limits the ability to represent non-verbal aspects of human interaction such as facial expressions, speech, gestures, and visual presence.

Recent research has therefore focused on improving conversational systems through Natural Language Processing, emotion and sentiment analysis, speech technologies, and intelligent virtual agents. Although these approaches can improve the understanding of user input and generate supportive responses, several systems still lack an integrated three-dimensional representation capable of responding through expressive facial and body animations. This creates an opportunity to combine

conversational AI with speech interaction, emotion-aware processing, and an interactive 3D avatar to provide a more engaging and human-like emotional-support experience. The proposed project addresses this direction by integrating conversational AI, emotion detection, speech recognition, speech synthesis, VRM-based avatar rendering, synchronized facial animation, dashboard visualization, and session reporting within a unified application.

2.1 Artificial Intelligence and Conversational Systems for Emotional Support

Artificial Intelligence (AI) provides an alternative to conventional rule-based support systems by enabling applications to process user input and generate context-sensitive responses. **Natural Language Processing, conversational AI, sentiment analysis, and emotion detection** have been increasingly incorporated into digital mental-health applications to understand users' messages and provide supportive interactions. Such systems can offer convenient and continuous communication, while conversational models can use recent dialogue context to produce more relevant responses. However, text-oriented systems may have limited emotional expressiveness because they cannot naturally represent non-verbal elements such as facial expressions, gestures, speech, and visual presence.

Recent developments in intelligent virtual agents extend conversational systems by combining **speech recognition, speech synthesis, emotion-aware processing, and three-dimensional avatar technologies.** These approaches can provide a more interactive form of communication by allowing users to communicate through both text and voice while presenting AI-generated responses through an expressive virtual character. The proposed project builds upon this direction by integrating conversational AI with **emotion detection and a VRM-based 3D avatar,** including facial expressions, lip synchronization, blinking, breathing, and body gestures. The system further incorporates dashboard-based emotional monitoring and automatic session-report generation to support visualization and review of interaction history.

Unlike conventional machine-learning approaches that depend heavily on manually selected features, the proposed system focuses on integrating multiple AI and human-computer interaction components into a unified emotional-support environment. However, the current emotion-detection component is based on **keyword-oriented analysis**, which may not capture complex emotional states in all conversational situations. Therefore, the system provides a foundation for future enhancement through more advanced emotion-recognition and contextual analysis techniques.

2.2 Conversational AI and Emotion-Aware Interaction

Conversational Artificial Intelligence (AI) enables computer systems to understand user messages and generate responses based on the information provided during an interaction. In emotional-support applications, conversational models can be used to maintain dialogue, interpret the context of recent messages, and provide supportive responses. Natural Language Processing (NLP) forms an important part of this process by allowing the system to process both typed input and text obtained from speech recognition. However, conversational interaction based only on textual responses may provide limited emotional expressiveness and may not reproduce important non-verbal aspects of human communication.

Emotion detection provides an additional layer by analysing the user's input and identifying an associated emotional state. In the proposed system, emotion detection is incorporated into the conversation pipeline so that the detected state can influence the response and the visual behaviour of the virtual avatar. The current implementation uses keyword-oriented analysis to identify emotions such as happiness, sadness, anxiety, anger, surprise, and neutral states. Although this approach is relatively simple to implement, it may not fully capture complex emotions, contextual meanings, or subtle changes in emotional state.

The combination of conversational processing and emotion-aware interaction can be represented as:

User Input → NLP Processing → Emotion Detection → AI Response Generation → Emotional Response → Avatar Interaction

This integration allows the proposed system to move beyond conventional text-based chatbot interaction by connecting conversational responses with an expressive three-dimensional avatar. The avatar can display corresponding facial expressions and other

animations, creating a more interactive form of communication.

2.3 Speech Interaction and 3D Virtual Avatar Technology

Speech-based interaction enables users to communicate with the system without relying exclusively on keyboard input. The proposed application supports voice interaction through browser-based speech recognition, where spoken input is converted into text before being processed by the conversational AI. The generated response is subsequently converted into speech using text-to-speech technology. This creates a bidirectional interaction process in which users can speak to the virtual assistant and receive an audible response.

Three-dimensional virtual avatars provide a visual representation of conversational systems and can improve the expressiveness of computer-mediated interaction. The proposed system uses a VRM-based avatar rendered through Three.js and incorporates facial expressions, synchronized lip movements, blinking, breathing, and body gestures. The emotional state detected from the user's input can influence the avatar's expression, allowing the visual response to correspond with the conversational context.

The overall interaction can therefore be expressed as:

Voice/Text Input → Speech/NLP Processing → Emotion Detection → Conversational AI → Text-to-Speech → 3D Avatar Animation

The combination of speech processing and avatar-based interaction provides a more immersive interface than a conventional text-only chatbot. Nevertheless, the quality of speech recognition, emotion detection, and avatar synchronization can affect the overall interaction experience. The proposed system therefore combines these components with dashboard monitoring and session management to provide a unified emotional-support environment.

2.4 Evaluation and Practical Monitoring

A conversational emotional-support system should be evaluated not only on its ability to generate responses but also on the functionality and reliability of its major interaction components. The proposed system therefore considers the performance of **text and voice communication, emotion detection, conversational response generation, avatar expressions, lip synchronization, dashboard updates, session management, and PDF report generation.** Functional, structural, and behavioral testing are used to examine whether individual modules and their interactions operate as intended. The testing process also considers system stability and usability during extended interactions.

In addition to conversational performance, practical interaction and monitoring are important for the proposed system. The application provides a dashboard that presents information such as the current emotional state, estimated stress level, mood trends, session duration, message count, and emotional progression. Session information, including conversation history and detected emotions, is maintained for subsequent analysis and report generation. The system also uses PDF generation to provide a structured summary of the completed interaction.

2.5 Research Gap

The reviewed literature indicates several limitations in existing AI-based emotional-support and mental-health chatbot systems. Many conventional systems rely primarily **on text-based communication**, which limits emotional expressiveness and does not adequately represent non-verbal aspects of human interaction. Existing approaches may also provide limited integration of speech interaction, facial expressions, animated avatars, emotional visualization, and long-term session reporting.

Another gap concerns the integration of multiple interaction technologies within a single system. While conversational AI can provide intelligent responses and emotion-analysis techniques can identify aspects of user sentiment, these capabilities are often considered separately from visual and speech-based interaction. The proposed project addresses this gap by combining **conversational AI, NLP, text and voice communication, emotion detection, speech recognition, speech synthesis, VRM-based 3D avatar rendering, synchronized facial and body animations, dashboard visualization, session management, and PDF report generation** within one unified framework.

The proposed system therefore focuses on improving **the interaction and visualization layer**

of AI-based emotional-support applications rather than claiming clinical replacement of professional therapy. Its current emotion-detection mechanism is keyword-oriented, and the project report identifies advanced emotion recognition, multilingual support, cloud storage, wearable integration, and professional monitoring as possible future enhancements.

3. RESEARCH METHODOLOGY

The proposed **Interactive 3D Digital Human for Real-Time Emotional Support and Therapy** follows a systematic approach for developing a multimodal AI-based emotional-support system. The methodology combines **Natural Language Processing, conversational AI, emotion detection, speech recognition, speech synthesis, and three-dimensional avatar interaction** within a unified web-based environment. The system accepts text or voice input from the user, processes the input to identify an associated emotional state, generates a context-aware response through the conversational AI component, and presents the response through both the interface and an expressive 3D avatar. Session information is simultaneously monitored and stored for dashboard visualization and report generation. The project report describes this as a modular architecture integrating AI, speech technologies, avatar rendering, emotion detection, visualization, and session reporting.

The overall workflow is:

User Text/Voice Input → Speech Recognition (if required) → NLP Processing → Emotion Detection → Conversational AI Response Generation → Text-to-Speech → 3D Avatar Animation → Dashboard Monitoring → Session Storage → PDF Report Generation

3.1 Research Design

The study adopts an **applied system-development and experimental evaluation approach**. The applied aspect focuses on developing a practical AI-based emotional-support application that combines conversational interaction with an expressive 3D digital human. The experimental aspect involves testing the functionality and interaction of the developed system, including text and voice communication, emotion detection, AI-generated responses, avatar animation, dashboard monitoring, session management, and report generation. The project report identifies static, structural, and behavioral testing as the major testing strategies used to evaluate the system.

The methodology consists of the following stages:

1. **User Input Processing**:
   The system accepts user input through either text or voice. When voice input is used, browser-based speech recognition converts the spoken content into text before further processing.

2. **Natural Language and Emotion Processing**:
   The received text is processed by the conversational pipeline, while the emotion-detection component analyzes the input to identify an associated emotional state. The current implementation uses keyword-oriented analysis for emotional categories such as happiness, sadness, anxiety, anger, surprise, and neutral states.

3. **Conversational AI Response Generation**:
   The processed user input and relevant conversational context are transferred to the AI processing component, which generates a supportive response. This enables the system to maintain continuity within the ongoing interaction.

4. **Speech Synthesis**:
   For voice-based interaction, the generated textual response is converted into speech using text-to-speech technology. This allows the virtual avatar to communicate the response audibly rather than relying only on displayed text.

5. **3D Avatar Interaction**:
   The response is presented through a VRM-based three-dimensional avatar rendered using Three.js. Avatar behavior includes facial expressions, synchronized lip movements, blinking, breathing, and body gestures. The detected emotional state can influence the avatar's visual expression during the interaction.

6. **Dashboard and Session Monitoring**:
   The system continuously updates session-related information, including current emotion, estimated stress level, mood

trend, session duration, message count, and emotional progression. This provides a visual representation of the user's interaction during a session.

7. **Session Storage and Report Generation:** Conversation history, detected emotions, completed exercises, session duration, and related user information are maintained within the browser environment using Local Storage. At the end of a session, the collected information is processed using jsPDF to generate a structured PDF report containing relevant conversation and emotional information.

This methodology provides an end-to-end framework connecting **multimodal user interaction, emotion-aware conversational AI, speech processing, 3D avatar animation, session monitoring, and automated reporting**. The modular structure also provides a foundation for future enhancements such as advanced emotion recognition, multilingual interaction, cloud-based storage, wearable integration, and professional monitoring.

3.2 Proposed System Modules

The proposed **AI-based emotional-support framework** consists of interconnected modules responsible for user interaction, conversational processing, emotion detection, speech communication, 3D avatar rendering, session monitoring, and report generation. The major modules are **User Login, Conversation Management, Voice Interaction, Emotion Detection, Conversational AI, Speech Synthesis, 3D Avatar Animation, Dashboard Monitoring, Session Management, Mood Tracking, and PDF Report Generation**.

These modules work together to provide a unified interaction in which user input is analyzed, an appropriate AI response is generated, and the response is presented through both the conversational interface and the expressive virtual avatar. The project follows a modular architecture in which each component performs a specific function while communicating with other components through defined interfaces.

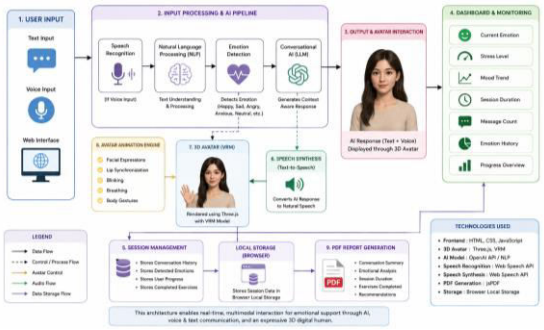

Fig 3.2 Proposed System Architecture

3.2.1 User Login Module

The **User Login Module** provides the initial access point to the application and allows users to enter the virtual emotional-support environment. The module manages the basic user-entry process and connects the user with the main conversational interface. After successful access, the user can interact with the AI system through text or voice and access the available session and monitoring features.

3.2.2 Conversation Management Module

The **Conversation Management Module** manages the exchange of messages between the user and the AI assistant. It accepts user messages, maintains the ongoing conversation context, and displays the generated responses within the interface. The module supports continuous interaction so that the AI response can consider recent conversational information rather than treating every message as an isolated request**.**

3.2.3 Voice Interaction Module

The **Voice Interaction Module** enables users to communicate with the system through spoken input. The browser-based speech-recognition functionality converts the user's speech into text, which is subsequently passed to the conversational processing pipeline. This module provides an alternative to keyboard-based interaction and supports more natural communication between the user and the virtual assistant.

3.2.4 Emotion Detection Module

The **Emotion Detection Module** analyzes the user's textual input to identify an associated emotional state. The current implementation uses keyword-oriented analysis to recognize emotional categories such **as happiness, sadness, anxiety, anger, surprise, and neutral**. The detected emotional state is used as additional information within the

interaction pipeline and can influence the visual behavior of the 3D avatar. The keyword-based approach provides a relatively simple mechanism for emotion identification, although it may not capture complex emotional expressions in every conversational context.

3.2.5 Natural Language Processing Module

The **Natural Language Processing (NLP)** Module processes the textual representation of the user's input and prepares it for conversational response generation. For voice interaction, the speech-recognition output is first converted into text before entering this stage. NLP enables the system to interpret the user's message and provide relevant information to the conversational AI component.

3.2.6 Conversational AI Module

The **Conversational AI Module** is responsible for generating responses to user messages. It receives the processed input and relevant conversational context and communicates with the AI service to obtain a context-aware response. The generated response is returned to the application and subsequently presented through text and, where applicable, speech and avatar interaction. The project architecture identifies conversational AI as a central component connecting user input with the response and avatar-output stages.

3.2.7 Speech Synthesis Module

The **Speech Synthesis Module** converts the AI-generated textual response into spoken audio. This allows the virtual assistant to respond through voice in addition to displaying the response as text. The generated audio also supports the avatar-animation process, particularly lip synchronization during speech.

3.2.8 3D Avatar Module

The **3D Avatar Module** provides the visual representation of the conversational AI. A VRM-based humanoid avatar is rendered using Three.js, allowing the user to interact with an expressive virtual character rather than a text-only interface. The avatar acts as the primary visual representation of the AI-generated response.

3.2.9 Avatar Animation Module

The **Avatar Animation Module** controls the non-verbal behavior of the virtual human. It incorporates **facial expressions, lip synchronization, blinking, breathing, and body gestures** to make the avatar's response more expressive. The detected emotional state can influence the avatar's facial expression, while speech output is associated with mouth movements during verbal responses.

3.2.10 Session Management Module

The **Session Management Module** manages information generated during an interaction session. It maintains conversation history, detected emotions, completed exercises, session duration, and other relevant session information. The stored information can subsequently be used for dashboard visualization and report generation.

3.2.11 Mood Tracking Module

The **Mood Tracking Module** monitors emotional information collected during the interaction and presents changes in the user's detected emotional state. The system can display information such as current emotion, estimated stress level, mood trends, and emotional progression. This provides a visual representation of session-level emotional information rather than displaying only individual chatbot responses.

3.2.12 Dashboard and Visualization Module

The **Dashboard and Visualization Module** provides a graphical overview of the ongoing or completed session. It presents information including **current emotional state, estimated stress level, mood trend, session duration, message count, and emotional progression**. The dashboard allows the collected session information to be viewed in an organized form and supports easier interpretation of interaction history.

3.2.13 Local Storage Module

The **Local Storage Module** maintains session-related information within the browser environment. The project uses browser Local Storage to retain conversation history, detected emotions, user information, session details, and completed exercises. This allows session information to remain available for dashboard presentation and report generation without requiring a separate database service.

3.2.14 Exercise and Support Module

The **Exercise and Support Module** provides supportive activities that can be presented during an interaction session. The project includes activities such as breathing and mindfulness exercises as part of the overall emotional-support experience. Completed activities can also form part of the session information maintained by the system.

3.2.15 PDF Report Generation Module

The **PDF Report Generation Module** converts collected session information into a structured report. The report can **include conversation history, emotional patterns, session duration, completed exercises, mood-related information, and recommendations.** The project uses **jsPDF** for generating the session report from the collected application data.

3.2.16 Monitoring and System Integration Module

The **Monitoring and System Integration Module** connects the individual components into a continuous interaction workflow. It coordinates user input, emotion analysis, conversational AI processing, speech output, avatar animation, session monitoring, and reporting. The integrated architecture enables the system to process a user's text or voice input, generate an AI response, express the response through the 3D avatar, and update the corresponding session information.

3.3 Input and Interaction Description

The proposed system supports **two primary forms of user input: text and voice.** Text input is directly submitted to the conversational processing pipeline, while voice input is first converted into text using speech-recognition functionality. The resulting textual representation is then processed for emotion detection and conversational response generation.

The overall interaction pipeline is:

Text/Voice Input → Speech Recognition (if required) → NLP Processing → Emotion Detection → Conversational AI → Response Generation

This multimodal input mechanism allows users to communicate with the system through either conventional text interaction or speech-based interaction.

3.4 Interaction Processing

The interaction-processing pipeline prepares user input for response generation through several connected stages:

User Input → Input Conversion → Text Processing → Emotion Detection → Contextual AI Processing → Response Generation

For voice interaction, the spoken input is converted into text before entering the processing pipeline. The resulting text is analyzed for emotional indicators and forwarded to the conversational AI component together with relevant conversational context. The generated response is subsequently delivered to the user through the interface and, when voice interaction is enabled, through speech synthesis.

3.5 Proposed Multimodal AI Architecture

The proposed architecture combines **conversational AI, emotion detection, speech processing, and 3D avatar technology** to create an interactive emotional-support environment. The overall structure is:

User Input → NLP & Emotion Detection → Conversational AI → Speech Synthesis / Text Response → 3D Avatar → Dashboard & Session Management

The system separates input processing from response presentation while maintaining communication between the modules. Emotion information can influence avatar expressions, while speech output supports lip synchronization and other animation behaviors. The architecture therefore combines linguistic, auditory, and visual interaction within a single application.

3.6 Emotion Detection Processing

The user's processed text is analyzed by the emotion-detection component to identify an associated emotional category. The current implementation uses keyword-based analysis rather than a deep-learning emotion-classification model. Identified categories include happiness, sadness, anxiety, anger, surprise, and neutral states.

The general process is:

User Text → Keyword Analysis → Emotion Category → Avatar/Response Context

The detected emotion provides additional information for the conversational interaction and can be used to determine an appropriate avatar expression. Since keyword-based analysis has limitations when dealing with ambiguous or context-dependent language, more advanced emotion-recognition techniques can be considered in future development.

3.7 Conversational AI Processing

After input and emotion processing, the information is provided to the conversational AI component. The AI model generates a response using the user's current message and available conversational context. The response is returned to the application and displayed to the user.

The response-generation flow is:

Processed Input → Conversational Context → AI Model → Generated Response → Text/Voice Output

This component forms the central intelligence layer of the proposed system and enables context-aware conversational interaction.

3.8 Avatar Response and Interaction

The generated response is presented through the **VRM-based 3D avatar**. The avatar is rendered using Three.js and supports multiple forms of animation, including facial expressions, blinking, breathing, body gestures, and lip synchronization.

Response Flow:

AI Response → Text-to-Speech → Audio + Emotion State → Avatar Animation → User

The resulting interaction combines the semantic response generated by the AI with visual and auditory feedback from the virtual human. This distinguishes the proposed system from a conventional text-only conversational interface.

3.9 Conversational Processing and Validation

The proposed system integrates conversational AI, emotion detection, speech processing, and 3D avatar interaction into a continuous processing pipeline. User input is first collected through text or voice interaction. Voice input is converted into text using speech-recognition functionality, after which the textual input is processed for emotional indicators and conversational context. The processed information is then provided to the conversational AI component to generate a response. The response is subsequently presented through text, speech, and the expressive 3D avatar.

Interaction Processing Workflow

Stage	Activity
1	Receive user text or voice input
2	Convert voice input into text when required
3	Process the textual input
4	Identify the associated emotional state
5	Provide input and conversational context to the AI component
6	Generate a context-aware response
7	Convert the response into speech when voice interaction is enabled
8	Determine appropriate avatar expression and animation
9	Present the response through the interface and 3D avatar
10	Update session information and dashboard

The validation process examines whether each stage operates correctly and whether the complete interaction can proceed from user input to AI response and avatar presentation.

3.10 System Architecture Description

The proposed system operates as an end-to-end multimodal interaction pipeline. User input is received through text or voice, processed using speech recognition and natural language processing when required, and analyzed for emotional indicators. The conversational AI component then generates a response based on the current input and available conversational context. The response is delivered through the user interface and can additionally be converted into speech and presented through the animated 3D avatar.

Session information generated during the interaction is used by the dashboard and reporting components. The system records information such as conversation history, detected emotions, session

duration, and completed activities, which can subsequently be used for visualization and PDF report generation.

3.11 Step-Wise System Architecture

The complete system follows these stages:

User Text/Voice Input → Speech Recognition → NLP Processing → Emotion Detection → Conversational AI → Response Generation → Speech Synthesis → 3D Avatar Animation → Dashboard → Session Storage → PDF Report

For text interaction, the speech-recognition stage is bypassed and the textual input is directly processed. For voice interaction, speech is first converted into text. The generated response can then be converted back into speech, while the avatar provides corresponding facial and body animations. The architecture therefore combines linguistic, auditory, and visual interaction within a unified application.

3.12 Tools and Technologies Used

The proposed system uses a combination of web technologies, artificial intelligence, speech technologies, visualization libraries, and 3D rendering technologies.

Technology	Purpose
HTML5	Structure of the web-based user interface
CSS3	Interface styling and layout
JavaScript	Client-side interaction and application logic
Node.js	Server-side application environment
Express.js	Backend/API development
Conversational AI / GPT	Context-aware response generation
Natural Language Processing	Processing and interpretation of user text
Web Speech API	Speech recognition and voice interaction
Text-to-Speech	Conversion of AI responses into spoken output
Three.js	3D rendering and avatar visualization
VRM / Three-VRM	Humanoid avatar representation and control
VRoid Studio	Creation/customization of the 3D avatar
Chart.js	Dashboard and mood-related visualization
Local Storage	Session and conversation information storage
jsPDF	Automatic generation of session reports

These technologies collectively support the complete workflow from user interaction and AI processing to avatar presentation, session monitoring, and reporting. The project report specifically identifies HTML/CSS/JavaScript, Node.js, Express.js, GPT-based conversational processing, Web Speech API, Three.js, VRM, Chart.js, Local Storage, and jsPDF within the implementation.

3.13 Web-Based Deployment Methodology

The conversational system is deployed as a web-based application in which the frontend provides the interaction environment and the backend manages communication with the AI processing service. The browser interface accepts user input and sends the relevant information to the backend, which processes the request and returns the generated response to the frontend.

Deployment Flow

User Input → Web Interface → Backend Processing → Conversational AI → Response → Frontend → Speech/Avatar → Session Monitoring

For text interaction, the user message is directly transmitted for processing. For voice interaction, speech is first converted into text before the request is processed. The resulting response is displayed in the interface and can be converted into speech for avatar-based interaction. The project implementation describes a Node.js/Express-based backend through which conversational requests are handled.

3.14 Session Management and Interaction History

The system maintains session-related information to support monitoring and later review of interactions. **Browser Local Storage** is used to retain information such as conversation history, detected emotional states, session details, user information, and completed exercises. This provides continuity for the session and allows information to be retrieved for dashboard visualization and report generation.

Unlike a system that only generates an immediate response, the proposed application therefore maintains interaction-related information that can be used to provide a broader view of the session.

3.15 System Implementation

The implementation covers the complete workflow from user interaction to session reporting:

User Access → Text/Voice Input → Speech Recognition → NLP Processing → Emotion Detection → Conversational AI → Response Generation → Speech Synthesis → 3D Avatar Animation → Session Monitoring → Local Storage → Dashboard → PDF Report → System Testing

The implementation combines individual technologies into a single application rather than treating conversational AI, emotion detection, speech interaction, and avatar animation as independent components. The project architecture specifically connects user input, emotion detection, conversational AI, speech processing, avatar animation, dashboard monitoring, and session reporting.

3.16 Validation

Validation is performed at two complementary levels**: component-level validation** and **system-level validation.** Component-level validation examines whether individual functions such as speech recognition, emotion detection, response generation, avatar animation, and report generation work correctly. System-level validation examines whether these components operate correctly when integrated into a complete user interaction.

Validation Level	Evaluation
Component Level	Text input, voice recognition, emotion detection, AI response generation, speech synthesis, avatar animation, dashboard updates, session storage, and PDF generation
Interaction Level	Text-to-response and voice-to-response workflows
Avatar Level	Facial expressions, blinking, breathing, gestures, and lip synchronization
Session Level	Conversation history, emotion tracking, session duration, and message count
System Level	End-to-end interaction, dashboard, session management, and report generation
Usability/Functional Level	User-visible behavior, interface response, stability, and successful completion of system operations

The project report describes static, structural, and behavioral testing approaches for evaluating both individual components and integrated system behavior.

3.17 Evaluation Criteria

The proposed system uses multiple functional and interaction-oriented criteria rather than relying on a single numerical measure.

Evaluation Criterion	Purpose
Text Interaction	Determines whether users can successfully send and receive conversational messages

Voice Recognition	Evaluates whether spoken input is correctly converted into text
AI Response Generation	Examines whether the system produces appropriate conversational responses
Emotion Detection	Evaluates whether predefined emotional categories are identified from user input
Speech Synthesis	Determines whether generated responses can be converted into speech
Avatar Expression	Evaluates whether the avatar displays the intended emotional expression
Lip Synchronization	Examines synchronization between speech output and mouth animation
Animation Behavior	Evaluates blinking, breathing, and body-gesture behavior
Dashboard Visualization	Determines whether session and emotional information is displayed correctly
Session Management	Evaluates storage and retrieval of conversation/session information
PDF Report Generation	Determines whether session information can be converted into a structured report
End-to-End Reliability	Evaluates whether the complete interaction pipeline operates successfully

These criteria provide a broader evaluation of the proposed system because the application is not solely a conversational model; it is an integrated multimodal human-computer interaction system.

3.18 Practical and Ethical Considerations

The proposed framework should be considered an **AI-based emotional-support system rather than a replacement for professional psychological or medical care.** The system is designed to facilitate conversational support and interactive engagement, but its responses and emotion detection should not be interpreted as clinical diagnosis or professional treatment. The project documentation similarly identifies the application as a supportive tool rather than a substitute for professional care.

The current emotion-detection mechanism is based on keyword-oriented analysis. Consequently, the system may not correctly identify complex, ambiguous, or context-dependent emotional states. Similarly, the quality of voice recognition can vary depending on speech conditions, while AI-generated responses depend on the underlying conversational model and its available context.

Practical deployment should therefore consider **emotion-detection limitations, response reliability, speech-recognition errors, privacy of stored session information, dependency on external AI services, browser compatibility, and responsible handling of sensitive conversations**. Since session information is stored using browser Local Storage, appropriate consideration should also be given to the security and privacy of information maintained on the user's device.

The project identifies several possible future improvements, including more advanced emotion recognition, multilingual interaction, cloud-based storage, facial-emotion analysis, wearable-device integration, and professional monitoring.

4. DATA ANALYSIS

4.1 Interaction and Session Data Analysis

The proposed **Interactive 3D Digital Human for Real-Time Emotional Support and Therapy** analyzes information generated during conversations between the user and the AI-based virtual human. Unlike a conventional dataset-based machine-learning system, the proposed application operates on **user-generated conversational and interaction data**, including text messages, speech-transcribed input, detected emotional states, session information, and generated responses.

The system processes user input through text or voice interaction. Voice input is first converted into text, after which the input is analyzed for emotional indicators and passed to the conversational AI component. The generated response is then presented through the interface and, when enabled, through speech and the 3D avatar. Session information is subsequently used for dashboard visualization and report generation.

Data Processing Flow:

User Text/Voice → Speech Recognition → NLP Processing → Emotion Detection → Conversational AI → Response Generation → Avatar/Speech Output → Session Storage → Dashboard/Report

For text-based interaction, the speech-recognition stage is not required. The resulting session information provides the basis for analyzing the behavior and functionality of the proposed system.

4.2 System Performance Analysis

The proposed system is analyzed according to its ability to provide continuous multimodal interaction and correctly coordinate its major functional components. The analysis considers conversational processing, emotion detection, voice communication, avatar behavior, session monitoring, and report generation.

Component	Function
Text Interaction	Accepts typed user messages and displays AI-generated responses
Voice Interaction	Converts spoken input into text for conversational processing
NLP Processing	Processes the textual representation of user input
Emotion Detection	Identifies an associated emotional category from user input
Conversational AI	Generates context-aware responses
Speech Synthesis	Converts generated responses into spoken output
3D Avatar	Provides the visual representation of the AI assistant
Avatar Animation	Controls expressions, blinking, breathing, gestures, and lip synchronization
Session Management	Maintains conversation and session-related information
Mood Tracking	Displays emotional state and mood-related trends
Dashboard	Presents session and emotional information visually
Local Storage	Stores conversation and session information in the browser
PDF Reporting	Generates structured reports from session information

The integration of these components enables the proposed application to function as a complete **multimodal emotional-support environment** rather than as a standalone conversational chatbot. The project architecture connects conversational AI, emotion detection, speech processing, avatar animation, dashboard monitoring, and session reporting.

4.3 Functional Performance Evaluation

The system is evaluated using multiple functional criteria because the effectiveness of the proposed application depends on the interaction of several independent components.

Evaluation Criterion	Purpose
Text Interaction	Determines whether messages can be successfully submitted and responses received
Voice Recognition	Evaluates conversion of spoken input into usable text
AI Response Generation	Examines whether the conversational component produces responses

Emotion Detection	Determines whether predefined emotional states can be identified
Speech Synthesis	Evaluates generation of audible responses
Avatar Expression	Determines whether the appropriate visual expression is displayed
Lip Synchronization	Examines synchronization between speech and mouth movement
Blinking and Breathing	Evaluates continuous avatar animation behavior
Dashboard Visualization	Determines whether session information is presented correctly
Mood Tracking	Evaluates representation of emotional progression
Session Management	Examines storage and retrieval of interaction information
PDF Report Generation	Determines whether session information can be converted into a report
End-to-End Interaction	Evaluates successful operation of the complete system workflow

The project report describes testing of text communication, voice communication, AI responses, emotion detection, facial expressions, lip synchronization, blinking, breathing, dashboard functionality, mood tracking, session management, and PDF generation.

4.4 System Analysis and Observations

The developed framework combines conversational artificial intelligence with speech interaction, emotion-aware processing, and an interactive 3D digital human. The major observations from the implemented system are:

- **Multimodal Interaction:** The system supports both text and voice-based communication.
- **Conversational Processing:** User input is processed by the conversational AI component to generate responses based on the available interaction context.
- **Emotion-Aware Interaction:** The system analyzes user input and identifies predefined emotional categories.
- **Speech-Based Response:** Generated responses can be converted into speech for voice-based interaction.
- **3D Embodiment:** A VRM-based humanoid avatar provides a visual representation of the AI assistant.
- **Expressive Animation:** The avatar supports facial expressions, blinking, breathing, gestures, and lip synchronization.
- **Mood Monitoring:** The dashboard presents current emotional state, estimated stress level, mood trends, session duration, message count, and emotional progression.
- **Session Management:** Conversation history, detected emotions, session information, and completed exercises can be maintained for later analysis.
- **Automated Reporting:** Session information can be converted into a structured PDF report using jsPDF.
- **Integrated Interaction:** The system combines conversational, auditory, and visual interaction instead of relying exclusively on text.

Overall, the proposed framework integrates **conversational AI, NLP, emotion detection, speech recognition, speech synthesis, 3D avatar animation, dashboard visualization, session management, and automated reporting** into a unified emotional-support application

5. IMPLEMENTATION, EXPERIMENTS AND RESULTS ANALYSIS

5.1 System Implementation

The proposed **Interactive 3D Digital Human for Real-Time Emotional Support and Therapy** is

implemented as an integrated web-based application that combines conversational Artificial Intelligence, Natural Language Processing, emotion detection, speech processing, three-dimensional avatar animation, dashboard monitoring, session management, and therapy report generation.

The system uses **HTML, CSS, and JavaScript** for the user interface and **Node.js with Express.js** for backend communication. **Three.js and Three-VRM** are used to render and control the three-dimensional VRM avatar, while the **Web Speech API** provides speech recognition and speech synthesis capabilities. The application communicates with an AI model through the backend to generate context-aware conversational responses.

The implementation begins when the user enters the system through the personalized login interface. The user can communicate with the virtual therapist using either text or voice input. Voice input is converted into text through speech recognition and forwarded for processing. The system then applies keyword-based emotion detection to identify emotional states such as happiness, sadness, anxiety, anger, or neutrality. The detected emotion is used to update the avatar's expression and dashboard information.

The conversational AI module receives the processed user input and generates an appropriate response while maintaining recent conversation context. The response can then be converted into speech using text-to-speech functionality. At the same time, the avatar performs facial expressions, blinking, breathing, gestures, and lip synchronization to create a more interactive experience.

The system also maintains session information using browser-based Local Storage. Conversation history, detected emotions, session information, and other relevant details are used to update the dashboard and generate a downloadable therapy session report using the **jsPDF** library.

Overall Implementation Flow

User Login → Text/Voice Input → Speech-to-Text (if required) → Emotion Detection → AI Response Generation → Text-to-Speech → 3D Avatar Animation → Dashboard Update → Session Storage → PDF Therapy Report

The implementation therefore combines multiple technologies into a single interactive platform rather than functioning as a conventional text-only chatbot. The major components include conversational AI, emotion detection, speech processing, VRM avatar animation, dashboard visualization, session management, and report generation.

5.2 Experimental Setup

The experimental evaluation of the proposed system focuses on verifying whether the major functional components operate correctly and whether the complete application provides a stable and interactive user experience.

Unlike a conventional machine-learning classification project, the present system is primarily evaluated through **functional, structural, behavioral, integration, and usability-oriented testing**. The evaluation therefore concentrates on the successful operation of system modules rather than reporting classification metrics such as ROC-AUC or F1-score.

The experimental setup includes the following major components:

- User login and personalized session initialization.
- Text-based communication with the virtual therapist.
- Voice-based communication using speech recognition.
- AI-generated conversational responses.
- Keyword-based emotion detection.
- Avatar facial-expression updates based on detected emotions.
- Lip synchronization, blinking, breathing, and body animations.
- Real-time dashboard updates.
- Session information management.
- PDF therapy-session report generation.
- Backend and frontend communication.

- Error handling when services such as the AI backend or microphone are unavailable.

The implementation uses a browser-based environment with HTML5, CSS3, JavaScript, Node.js, Express.js, Three.js, Three-VRM, speech APIs, and AI services. The documented system requirements specify a minimum environment of approximately 8 GB RAM, an Intel Core i5/AMD Ryzen 5 or higher processor, suitable graphics capability, microphone, audio output, and an internet connection for AI services.

The testing process evaluates both individual modules and the complete integrated system. Particular attention is given to the interaction between the frontend, backend, AI service, speech APIs, emotion detector, avatar controller, dashboard, and report-generation module.

5.3 Experimental Results

The experimental results demonstrate the functional behaviour of the developed virtual therapy system. Testing was performed using different user inputs and operating conditions to determine whether the expected system behaviour was achieved.

The evaluation considered user login, AI conversation, emotion detection, backend failure handling, and speech recognition under adverse conditions. The documented test cases produced the following results:

Test Case	Evaluation Area	Result	Status
Test Case 1	User login and session initialization	Dashboard opened successfully and user information was displayed correctly	**Pass**
Test Case 2	AI conversation and emotion detection	Emotion was detected, an AI response was generated, and the avatar updated successfully	**Pass**
Test Case 3	AI backend unavailable	System displayed an error message, but conversation could not continue until the backend was restarted	**Fail**
Test Case 4	Speech recognition under poor audio conditions	Recognition failed because of environmental noise or microphone-permission restrictions	**Fail**

The first test case confirmed that the login functionality operated correctly. After valid user and therapist information was entered, the application successfully opened the therapy dashboard and displayed the user's information.

The second test case evaluated the core interaction between emotion detection, conversational AI, dashboard updates, and avatar behaviour. When the input indicated that the user was feeling stressed, the system detected the corresponding emotional state, generated an AI response, and successfully updated the avatar. This confirms the integration between the emotion-detection and conversational components.

The third test case examined system behaviour when the AI backend was unavailable. The application displayed the message **"Unable to connect to AI."** and did not crash; however, the conversation could not continue until the backend was restarted. Therefore, this test case was recorded as a failure and indicates an area where improved offline or fallback handling could be introduced.

The fourth test case evaluated speech recognition under poor audio conditions or when microphone permission was unavailable. The system was unable to reliably recognize voice input in these circumstances, resulting in a failed test case. This indicates that speech interaction remains dependent on microphone availability, browser permissions, and environmental audio quality.

Overall, the results show that the major interactive functions of the proposed system operate

successfully under normal conditions. The system is capable of accepting user input, detecting emotions, generating AI responses, controlling the 3D avatar, and providing an interactive therapy environment. At the same time, the failed backend and speech-recognition test cases reveal practical limitations that should be addressed in future versions.

The testing process therefore demonstrates the feasibility of integrating **AI-based conversation, emotion detection, speech interaction, VRM avatar animation, dashboard monitoring, and therapy-session reporting** into a unified web application. The results also provide useful evidence for further improvement in error handling, speech

5.4 Graphical Analysis

Graphical analysis is used to visually represent the behaviour and performance of the proposed **Interactive 3D Digital Human for Real-Time Emotional Support and Therapy**. The system collects interaction-related information during a therapy session, including detected emotional states, stress indicators, session duration, message count, and mood progression. These values can be presented through dashboard graphs to make the user's interaction and emotional changes easier to understand.

The dashboard provides visual indicators for the current emotional state and estimated stress level. It also maintains mood trends and emotional progression throughout the session. These graphical elements allow users to observe how their detected emotional state changes during interaction with the virtual therapist.

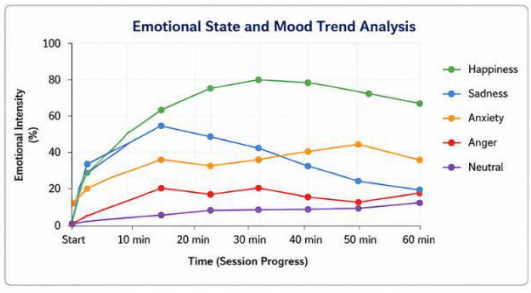

Fig. 5.1. Emotional State and Mood Trend Analysis

The mood-trend graph represents changes in the emotional state detected from user messages during a therapy session. Since the current implementation uses **keyword-based emotion analysis**, the detected categories include emotions such as happiness, sadness, anxiety, anger, surprise, and neutrality.

The graphical representation provides a simple way to observe variations in emotional states throughout the conversation. A change in the detected emotional category may indicate a corresponding change in the content of the user's interaction. However, these results should be interpreted as **system-detected emotional categories rather than clinical measurements**.

Dashboard Analysis

The dashboard also presents additional session-related information, including:

- Current detected emotion
- Estimated stress level
- Mood trend
- Session duration
- Number of messages
- Emotional progression

These parameters provide a summarized view of the ongoing session and can later be incorporated into the generated therapy report.

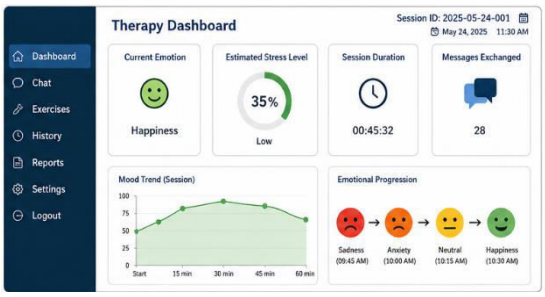

Fig. 5.2. Dashboard Monitoring and Session Statistics

The dashboard visualization improves the readability of session information by presenting important parameters in an organized graphical form. It also enables users to review their interaction patterns without manually examining the complete conversation history.

Emotion Distribution Analysis

An emotion-distribution graph can be used to represent the number of detected instances belonging to each supported emotional category during a session.

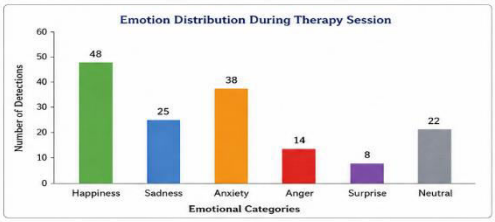

Fig. 5.3. Emotion Distribution During Therapy Session

The distribution provides an overview of which emotional categories were detected most frequently during the interaction. For example, if sadness or anxiety occurs more frequently than other categories, the graph will reflect a higher number of detections for those categories.

Because the current emotion detector is based on keywords, the distribution represents **detected textual emotional cues** rather than a comprehensive measurement of the user's psychological state. The project report itself identifies keyword-based emotion detection as the current approach and suggests advanced emotion-recognition methods as future enhancements.

System Testing Analysis

In addition to dashboard graphs, the system's functional test results can be represented graphically. Four test cases were documented in the project report. Two test cases passed and two failed.

Test Case	Result
User Login	Pass
AI Conversation and Emotion Detection	Pass
AI Backend Unavailable	Fail
Speech Recognition Under Poor Conditions	Fail

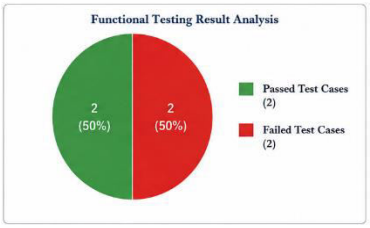

Fig. 5.4. Functional Testing Result Analysis

The test-result graph can display the number of successful and unsuccessful test cases. The results show that the system successfully performed user login and the normal AI conversation/emotion-detection workflow, while limitations were observed during backend unavailability and poor speech-input conditions.

5.5 Results Analysis

The experimental evaluation assessed the functionality and reliability of the proposed **Interactive 3D Digital Human for Real-Time Emotional Support and Therapy**. Since the system integrates conversational AI, emotion detection, speech processing, avatar animation, dashboard monitoring, and session management, traditional ML metrics such as accuracy and F1-score were not used.

The main workflow is:

User Input → Emotion Detection → AI Response → Speech Output → Avatar Animation → Dashboard Update → Session Storage → Report

Functional Results: User login, AI conversation, emotion detection, avatar updates, dashboard monitoring, session management, and report generation worked successfully under normal conditions. However, the system could not continue conversation when the AI backend was unavailable, and speech recognition was unreliable under poor audio or restricted microphone conditions.

Avatar and Dashboard Results: The 3D avatar supports facial expressions, blinking, breathing, gestures, and lip synchronization. The dashboard displays detected emotion, stress level, mood trends, session duration, and message count. Session data is stored in Local Storage and used for report generation.

Overall, the results demonstrate the feasibility of combining conversational AI with an interactive 3D

avatar for emotional-support applications. However, keyword-based emotion detection should not be considered a clinical diagnosis. Future improvements may include advanced emotion recognition, better speech processing, multilingual support, and cloud-based storage.

6. CONCLUSION AND FUTURE WORK

The proposed **Interactive 3D Digital Human for Real-Time Emotional Support and Therapy** successfully combines **Artificial Intelligence, NLP, emotion detection, speech processing, and 3D avatar technology** to provide an interactive emotional-support experience. The system supports text and voice communication, detects emotions, generates AI responses, and displays realistic avatar expressions and animations.

The system also provides **mood tracking, stress-level estimation, conversation history, and therapy-session report generation**, making it more interactive than a conventional text-based chatbot. However, it is intended as a supportive system and **not as a replacement for professional psychological counselling**.

6.1 Future Work

Future improvements can include:

- **Advanced emotion detection** using facial expressions, voice, and physiological signals.
- **Improved AI responses** using advanced conversational models.
- **More realistic avatar animation** with better facial expressions and gestures.
- **VR/AR integration** for immersive interaction.
- **Multilingual support** for wider accessibility.
- **Cloud deployment** for better scalability and accessibility.
- **Wearable-device integration** for additional physiological information.
- **Improved speech recognition and error handling** for greater reliability.

Overall, these enhancements can make the proposed system **more intelligent, realistic, accessible, and reliable** for future emotional-support applications.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–444, 2015.

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[4] R. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[5] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[6] M. Pantic and L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.

[7] A. H. B. Zadeh, M. Soleymani, A. S. Morency, and M. Pantic, "Multimodal Sentiment and Emotion Recognition in Text and Speech," *Proceedings of the IEEE*, relevant work on multimodal affective computing.

[8] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[9] A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[10] A. Radford et al., "Language Models are Unsupervised Multitask Learners," OpenAI, 2019.

[11] M. McTear, *Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots*. Cham, Switzerland: Springer, 2021.

[12] J. K. Burgoon, L. K. Guerrero, and V. Manusov, *Nonverbal Communication*. Routledge, 2016.

[13] Khronos Group, *glTF 2.0 Specification*, Khronos Group.

[14] Three.js, *Three.js Documentation*, Three.js.

[15] pixiv, *VRM Documentation and VRM Specification*, VRM Consortium.