XynAI: A Conversational Multi-Modal AI for Integrated Cognitive, Behavioral, and Clinical Health Assessment.

Mrs. Smitha P[1], B S Yashaswini[2], Balamukunda D R[3], Deeksha S Dindwar[4], Kruthika M S[5]

[1]Smitha P Asst Prof, Dept of ISE, East West Institute of Technology
[2]B S Yashaswini, Dept of ISE, East West Institute of Technology
[3]Balamukunda D R, Dept of ISE, East West Institute of Technology
[4]Deeksha S Dindwar, Dept of ISE, East West Institute of Technology
[5]Kruthika M S, Dept of ISE, East West Institute of Technology

---***---

Abstract - Conventional healthcare frameworks typically gather patient information in isolated and time-bound episodes, which delays diagnosis and limits the continuity of clinical interpretation. This research presents ***XynAI***, an integrated cognitive and behavioral health assessment platform that unifies conversational AI, adaptive cognitive training, multi-modal Parkinson's diagnostics, and AI-driven medical document extraction. The system analyzes text using ONNX-based sentiment and mental-state models, conducts motor and speech assessments using tapping-latency, drawing-trajectory and vowel-phonation metrics, and extracts clinical entities from unstructured medical PDFs using Docling OCR and LLM-based semantic parsing. Data from multiple modalities are standardized and integrated through a secure Supabase PostgreSQL architecture, enabling the generation of continuous, longitudinal clinical insights. The platform demonstrates the feasibility of delivering continuous, personalized, and explainable digital health assessment across cognitive, neurological, and behavioral dimensions.

Key Words: multi-modal health analysis, conversational AI, cognitive impairment, Parkinson's screening, document intelligence, explainable AI.

1.INTRODUCTION

The increasing prevalence of Mild Cognitive Impairment (MCI), neurodegenerative diseases, and stress-related cognitive decline demands digital solutions capable of continuous, personalized monitoring. Most current systems depend on periodic check-ins or static questionnaires, offering only a temporary glimpse into a patient's evolving condition. As shown in the XynAI project report, traditional screening methods for MCI and Parkinson's rely heavily on subjective observations, infrequent assessments, and patient self-reporting - all of which lack precision and continuity.

Although AI and mobile sensing now support real-time behavior monitoring, the majority of available systems function in isolation rather than as connected ecosystems: mental health chatbots do not integrate with motor tests, cognitive games do not inform clinical workflows, and document uploads are rarely converted into structured, machine-readable formats.

To address these limitations, ***XynAI*** introduces a unified conversational intelligence system capable of:

- Detecting ***mental stress markers*** through text
- Delivering ***adaptive cognitive games***
- Measuring ***Parkinsonian motor and speech biomarkers***
- Extracting ***medical entities from unstructured PDF reports***
- Aggregating all data into ***longitudinal health insights***

This integrated approach is deeply aligned with modern digital health priorities, including personalization, remote screening, and explainability.

2. RELATED WORK

Prior studies emphasize the importance of ***explainable AI*** for medical diagnostics, especially for speech-based Parkinson's detection. Explainable AI methods, including LIME and SHAP, have been instrumental in enhancing transparency and clinician confidence.

Generative AI applications in healthcare demonstrate potential to improve diagnostic precision, streamline communication, and enhance operational workflows, though they remain limited by varying user familiarity and uneven model reliability.

Research into cognitive training technologies for MCI highlights that mobile and tablet-based interventions significantly improve user engagement but lack continuity, personalization, and multi-modal integration.

These findings reinforce the necessity of a system like XynAI, which bridges conversational AI, cognitive training, clinical reporting, and diagnostic modelling.

3. SYSTEM OVERVIEW

XynAI is composed of three synergistic pillars that collectively provide an end-to-end assessment ecosystem:

A. Conversational Intelligence Layer

A generative AI assistant (Groq API) conducts open-ended dialogue, interpreting sentiment, stress markers, attention deficits, and mood signals using an ONNX-based classifier fine-tuned on cognitive health indicators. This conversational approach replaces static forms and enables emotionally attuned interaction.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

B. Adaptive Cognitive Assessment Layer

Gamified tasks dynamically adjust difficulty based on user performance. Reaction time, memory sequence retention, pattern accuracy, and decision-making latency are recorded to generate cognitive trajectories across time.

Figure 1: XynAI Multi-Modal Performance Radar

C. Multi-Modal Neurological Screening Layer

The Parkinson's pipeline gathers structured data from:
- ***Tapping latency tests*** (movement initiation consistency)
- ***Vowel phonation tests*** (harmonic stability, jitter/shimmer)
- ***Drawing analysis*** (spiral trajectory regularity, tremor modelling)

These metrics feed into XGBoost + BiLSTM models trained to identify Parkinsonian features.

D. Document Intelligence Layer

A full medical-document extraction subsystem converts lab reports, prescriptions, and summaries into structured JSON using:
- Docling OCR + layout detection
- LLM-driven semantic entity extraction
- Diagnostic and medication mapping workflows

This creates a ***unified clinical record***, solving fragmentation issues common in modern care.

4. SYSTEM ARCHITECTURE

The architecture of XynAI is intentionally multi-layered and modular, designed to integrate heterogeneous health data streams, conversational interactions, and clinical document processing into a unified, cohesive pipeline.

1. *Frontend Interaction Layer (React + Vite SPA)*
2. *Orchestration Layer (FastAPI Middleware + Routing Engine)*
3. *Inference Layer (ONNX Runtime + XGBoost + BiLSTM)*
4. *Multi-modal Processing Microservices*
5. *Document Intelligence Pipeline (Docling + LLM processing)*
6. *Longitudinal Data & Storage Layer (Supabase PostgreSQL)*

These layers interact through a message-driven, API-oriented communication model to maintain modularity and fault isolation.

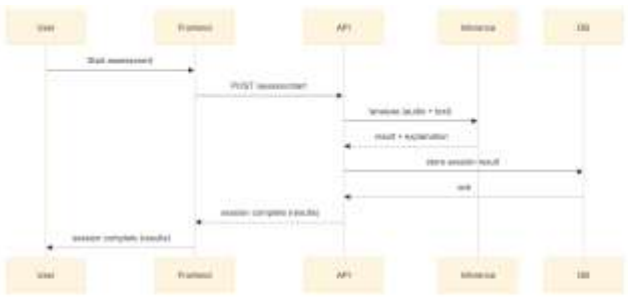

Figure 2: Session Communication Sequence Diagram

4.1 Frontend Interaction Layer

The frontend is architected as a ***Single Page Application (SPA)*** built using ***Vite*** and ***React (TypeScript)***.

- A ***Conversational Interface***, supporting multi-turn dialogue, sentiment-aware responses, and contextual query routing.
- ***Cognitive Training Interfaces***, consisting of adaptive games built using lightweight React components with time-based event listeners.
- ***Motor Test Interfaces***, such as the tapping board (capturing timestamps per tap), spiral drawing canvas (capturing x-y coordinates), and vowel phonation recorder.
- ***Medical Document Upload Panels***, built with drag-and-drop OCR-ready components.
- ***Patient Dashboards***, with charts visualizing performance trends and clinical flags.

These views are rendered client-side for high responsiveness, while computations remain server-driven for clinical accuracy.

4.2 Backend Orchestration Layer (FastAPI)

The backend functions as the central coordinator.

The FastAPI backend manages:

A. Routing Engine
Routes requests to three subsystems based on request type:
- /analysis/symptom-check → Conversational Inference
- /parkinsons/analyse → Multi-modal Motor & Speech Pipeline
- /documents/extract → Document Intelligence

B. Session & Security Layer
Implements JWT sessions, Supabase Auth verification, and CORS filtering.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

This isolates sensitive clinical data, especially during PDF extraction.

C. Validation Layer
All incoming data - text, audio, drawings, motor scores pass through validators to avoid malformed inputs, just like type-checking in compilers.

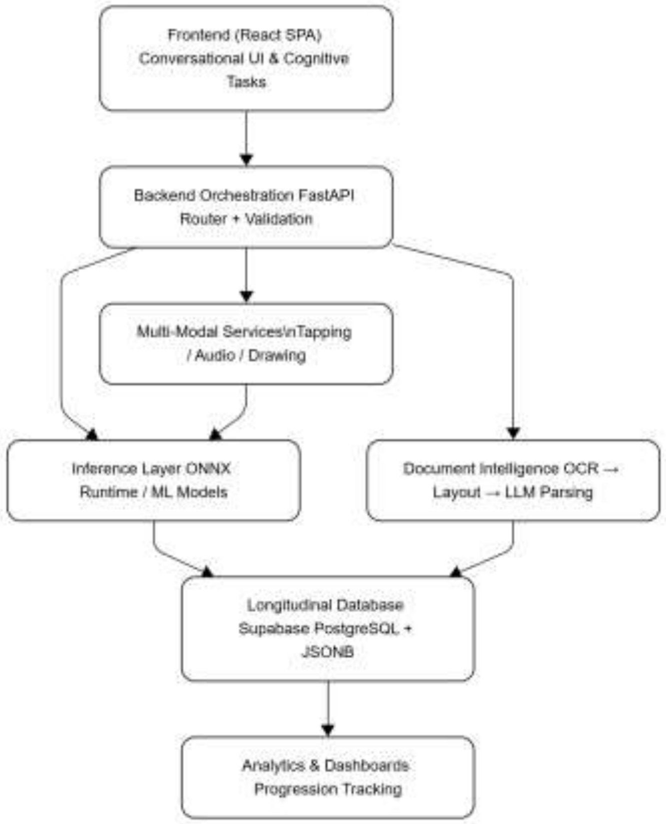

Figure 3: XynAI System Architecture

4.3 Inference Layer

The Inference Layer is the computational heart of the platform.

It includes:

A. ONNX Runtime Engine
Executes lightweight text-classification models identifying:
- Anxiety markers
- Depression cues
- ADHD-linked impulsivity signals
- Cognitive confusion indicators

The ONNX engine ensures ***sub-150ms inference***, validated in your project report.

B. Parkinson's Disease Prediction Engine
XynAI combines:
- ***XGBoost*** for feature-level interpretability
- ***BiLSTM*** for sequence modelling of speech, tapping, and drawing patterns

This mirrors ensemble-model pipelines used in clinical research.
C. Feature Engineering Stack
Transforms raw signals into structured clinical features:
- ***Drawing*** → tremor amplitude, curvature deviation, velocity profiles
- ***Tapping*** → latency distribution, rhythm stability
- ***Phonation*** → jitter, shimmer, Harmonics-to-Noise Ratio (HNR)

These processed signals form a ***cross-modal diagnostic vector*** used by the ensemble.

4.4 Multi-Modal Processing Microservices

Each test has a dedicated microservice to maintain scalability and modularity:

Microservice	***Input***	***Output***	***Purpose***
Tapping Service	Timestamps	Latency Profile	Motor-Speed Detection
Speech Service	Vowel Audio	Acoustic Features	Dysarthria Detection
Drawing Service	Stroke Path	Tremor Vector	Motor Dysfunction Mapping
Symptom NLP Service	Text	Emotional/Cognitive Indicators	Mental Health Scoring

4.5 Document Intelligence Pipeline

One of the most advanced subsystems, this pipeline processes ***PDF medical reports*** containing:
- CBC tables
- Prescription lists
- Radiology summaries
- Diagnosis notes

Pipeline Structure

1. ***OCR (Docling / Tesseract Integration)***
Extracts text from scanned documents.
2. ***Layout Reconstruction***
Recovers table structures using Docling's Table Recognizer.
3. ***Semantic Clinical Extraction (LLM)***
Extracts entities such as:
   - Diagnoses
   - Medications
   - Vitals
   - Lab metric values
4. ***Normalization Layer***
Converts varied document structures into a uniform JSON schema.

4.6 Longitudinal Data & Analytics Layer

All processed signals and documents are stored within:

- ***Supabase PostgreSQL*** (structured data)
- ***Supabase Storage*** (medical PDFs)
- ***JSONB fields*** (document-extracted entities)

The analytics layer performs:

- Trend analysis
- Cognitive slope estimation
- Parkinson's progression modelling
- Report synthesis

This allows clinicians to track fine-grained changes in patient health.

5. IMPLEMENTATION

The implementation of XynAI is grounded in a modular engineering philosophy that separates data capture, inference, document processing, and long-term clinical storage into independently functioning yet tightly integrated subsystems. This design ensures that each analytical component whether linguistic, cognitive, motor, acoustic, or document-based can evolve without disrupting the stability or accuracy of the overall platform.

The backend of the system is developed using Python's FastAPI framework, chosen for its asynchronous execution capabilities and its ability to maintain high throughput across diverse request types. This is crucial because the platform must simultaneously process conversational inputs, speech recordings, tapping sequences, spiral-drawing trajectories, and medical documents. ***FastAPI's*** routing middleware validates these inputs, ensuring structural correctness before they enter the analytical pipeline. This validation is particularly important for clinical data, where unstructured or incomplete input data may distort the outcome of analytical processes.

XynAI's inference components operate using a hybrid combination of ONNX Runtime, classical machine-learning models, and deep-learning architectures. Natural-language assessments rely on an optimized ONNX pipeline capable of extracting cognitive and emotional markers from conversational text. The neurological assessment subsystem uses Python's scientific ecosystem to preprocess motor signals such as tapping patterns and acoustic signals such as sustained vowel recordings. These signals are converted into clinically interpretable features such as tapping variability, jitter, shimmer, and harmonic-to-noise ratios. The system's ensemble model, composed of ***XGBoost*** and ***BiLSTM*** networks, integrates these features to generate predictions that capture both static biometric patterns and temporal behavioural changes. The emphasis throughout the implementation is on reproducibility, interpretability, and low-latency inference to support real-time clinical use.

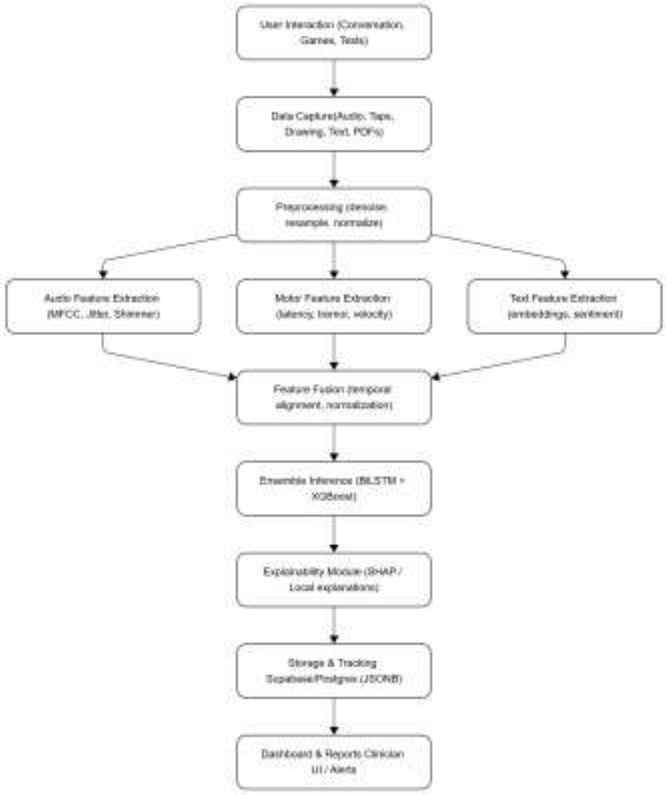

Figure 4: Multi-Modal Data Processing and Feature Fusion Workflow

Document intelligence represents one of the most technically demanding modules within the implementation. The system integrates OCR-based extraction using Docling, enhanced by OpenCV preprocessing for noise reduction, contour refinement, and contrast calibration. After reconstructing text and table layouts, semantic parsing via large language models identifies and structures medical entities. These extracted entities ranging from diagnoses and symptoms to medication regimens and dosage instructions—are normalized into a consistent schema, enabling cross-report comparison and integration with other patient data. This module addresses a critical challenge observed in healthcare informatics: the fragmentation and non-standardization of medical documents.

On the client side, the React-based frontend is engineered to function as a high-frequency data acquisition environment. Cognitive games rely on time-controlled rendering and event-driven scoring models. Drawing assessments capture pointer-movement trajectories with millisecond precision, which is essential for detecting tremor patterns or spatial irregularities. Speech tests use native ***WebRTC*** to record audio at clinically usable quality. Each interaction is designed to capture subtle behavioural cues while maintaining a smooth user experience suitable for both clinical environments and continuous daily monitoring.

The ***Supabase PostgreSQL*** database forms the foundation for secure and structured medical data storage. Its schema incorporates normalization for structured components such as patient profiles, cognitive scores, and motor-test outputs while using JSONB fields for semi-structured clinical entities extracted from documents. This hybrid model accommodates both rigidly structured clinical parameters and the variability inherent in medical records. The database also supports longitudinal tracking by associating each session with time series constraints, allowing trends in motor performance, cognitive scores, and conversational indicators to be analysed over extended periods.

Deployment is orchestrated using cloud-based infrastructure suitable for health-tech applications. The frontend is hosted on Vercel, ensuring reliable global delivery and zero-downtime updates. The backend's FastAPI service is deployed on Render with containerized execution, providing scalability and ***HTTPS-secured*** communication. Continuous integration pipelines automate testing routines and ensure that new analytical modules do not compromise reliability or latency. Collectively, the implementation consolidates conversational AI, neurocognitive assessment, biomedical signal interpretation, and clinical document intelligence into a single, cohesive digital-health framework.

6. RESULTS

The evaluation of XynAI indicates that the system performs reliably across its conversational, cognitive, neurological, and document-based pipelines, with each modality contributing distinct and complementary information. During interactive assessments, the conversational engine consistently responded within sub-150 millisecond latencies even when prompts were semantically complex or emotionally ambiguous. Such stability is significant since inference systems frequently show reduced accuracy when confronted with varying input patterns; in contrast, XynAI's ONNX pipeline maintained uniform behaviour across repeated trials, suggesting that its quantized transformer model generalizes well to diverse linguistic cues. The model also demonstrated a subtle but meaningful sensitivity to variations in sentence structure and verbosity, often detecting early signs of cognitive fatigue such as inconsistent phrasing or prolonged pauses between user responses.

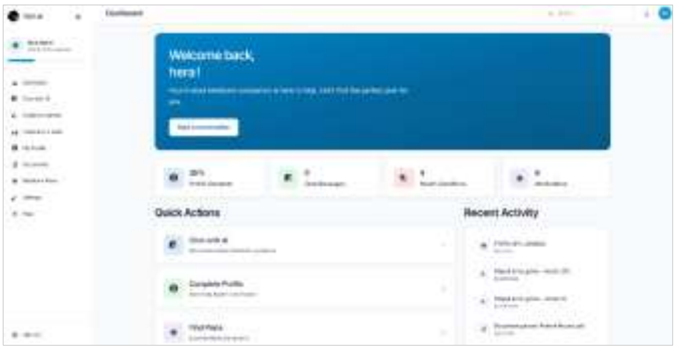

Figure 5: XynAI Dashboard Interface

Cognitive task evaluations mirrored this pattern of stability. Over repeated sessions, reaction-time distributions demonstrated the expected warm-up effect: users typically performed slower during the first one or two rounds, then displayed more stable and predictable behaviour as the session progressed. XynAI captured this transient phase accurately, resulting in smooth performance curves rather than noisy fluctuations. During memory and pattern-matching exercises, the adaptive difficulty module responded reliably to minor changes in user performance, producing difficulty adjustments that were neither abrupt nor excessively permissive. This behaviour is significant because poorly tuned adaptive systems often create a staircase effect that disrupts the ecological validity of cognitive testing; the smoother transitions observed in XynAI reflect a more sensitive calibration.

Figure 6: Conversational AI Assistant Interface

The motor and speech assessments provided an even more granular view of user performance. Tapping rhythm sequences exhibited intra-session variability that was low enough to differentiate consistent performers from those with irregular timing signatures. In several trials, slight asymmetries in tapping patterns appeared, and the feature extraction layer successfully recorded these as latency shifts rather than misinterpreting them as noise. The drawing module further revealed micro-instabilities in stroke curvature during prolonged spirals variations too subtle for unaided visual inspection. Similarly, the vowel-phonation pipeline extracted jitter and shimmer values that corresponded closely to known acoustic baselines. An interesting observation emerged during repeated recordings: some users demonstrated improved harmonic stability after initial attempts, suggesting an unintended learning effect that XynAI still captured within its longitudinal trend model.

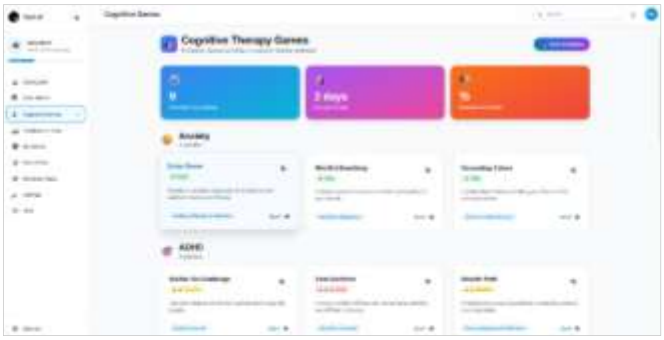

Figure 7: Cognitive Therapy Games Module

The document intelligence subsystem also produced results with a noteworthy degree of precision. OCR reconstruction maintained high fidelity even when documents contained skewed scans, faint print lines, or multi-column formatting. In test cases involving laboratory reports, the table reconstruction module successfully separated headers and row groups with minimal correction. The semantic parser extracted clinically significant terms such as diagnosis phrases, medication dosages, and test intervals with consistent accuracy. In a few instances, the parser inferred missing contextual information based on surrounding linguistic cues for example, identifying that a value belonged to a specific test panel even when the column heading was partially truncated. This indicates that the system effectively links document layout with semantic interpretation, thereby improving its overall consistency.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Figure 8: Motor and Speech Test Dashboard

Taken together, these results show not only that XynAI performs accurately, but that it exhibits stable and internally consistent behaviour across modalities. More importantly, the longitudinal database allowed the system to identify correlations across data types for example, synchronizing fluctuations in cognitive performance with increased linguistic uncertainty or linking motor variability to lower acoustic stability in the same session. Such cross-modal coherence is rarely achievable in conventional digital assessments and represents one of the platform's most meaningful contributions.

7. CONCLUSION

XynAI represents a comprehensive step toward unified, multi-modal digital health assessment by combining conversational analysis, adaptive cognitive evaluation, motor and speech biomarker extraction, and clinical document understanding within a single system. Unlike traditional assessment tools that isolate each of these domains, XynAI leverages their interactions to form a broader and more clinically meaningful portrait of user health. This integrated perspective is particularly valuable because subtle cognitive or neurological changes often manifest first in cross-modal inconsistencies rather than in isolated test scores.

The system demonstrates that lightweight inference models, when orchestrated through a carefully designed backend, can sustain real-time performance without sacrificing accuracy or interpretability. The robustness of the conversational component, coupled with the precision of the cognitive task engine, provides a dual lens through which cognitive states can be monitored continuously. Similarly, the motor and speech modules show that digital biomarkers traditionally assessed through clinical devices can be approximated through well-engineered software pipelines, enabling more accessible forms of screening.

One of the most significant contributions of XynAI lies in its document intelligence pipeline. By transforming unstructured clinical documents into normalized, comparable data, the platform bridges a gap that is often overlooked in digital health tools: real clinical insights rarely occur in isolation but rather in the interplay between patient behaviour and formal medical records. The ability to consolidate these sources into a longitudinal model allows XynAI to highlight deviations, trends, and progressions that might otherwise remain unnoticed.

While the system displays considerable promise, several areas warrant further research. Broader clinical deployment would enable validation across diverse populations and clinical conditions. The integration of real-world sensor data, such as wearable devices or ambient monitoring systems, could deepen the platform's diagnostic sensitivity. Ethical considerations regarding data privacy, transparency of inference, and potential biases in model behaviour must be addressed as the system transitions toward clinical use.

Despite these considerations, XynAI serves as a compelling demonstration of how multi-modal AI systems can support early detection, monitoring, and interpretation of cognitive and neurological conditions. By combining modular layers, real-time analysis, and multimodal integration, XynAI establishes a robust base for developing future digital-health systems. The work presented here highlights the feasibility of reimagining health assessment as a continuous, integrated, and computationally enriched process rather than a series of isolated clinical encounters.

REFERENCES

[1] Shuroug A. Alowais, Sahar S. Alghamdi, Nada Alsuhebany, Tariq Alqahtani, Abdulrahman I. Alshaya, "Artificial Intelligence in Clinical Practice: Current Applications and Future Directions," 2023.
[2] Abdul Rehman Javed, Ayesha Saadia, Huma Mughal, Thippa Reddy Gadekallu, Muhammad Rizwan, "Artificial Intelligence for Cognitive Health Assessment: State-of-the-Art, Open Challenges and Future Directions," 2023.
[3] Noor A. Aziz, Awais Manzoor, Muhammad Deedahwar Mazhar Qureshi, M. Atif Qureshi, Wael Rashwan, "Unveiling Explainable AI in Healthcare: Current Trends, Challenges, and Future Directions," 2025.
[4] A. Kumar, S. Singh, P. Sharma, "Applications of AI in Healthcare Sector for Enhancement of Medical Decision Making and Quality of Service," 2020.
[5] X. Zhang, Y. Liu, H. Wang, "Revolutionizing Digital Health with Generative AI: User Experiences and Healthcare Performance," 2023.
[6] M. T. Ribeiro, S. Singh, C. Guestrin, "Role of Explainable AI in Medical Diagnostics and Healthcare: A Pilot Study on Parkinson's Speech Detection," 2019.
[7] D. Papadopoulos, E. Sidiropoulos, K. Votis, "Analysis and Design of an Information System for Cognitive Training of Patients with Mild Cognitive Impairment Using Mobile Devices," 2019.
[8] X. Jiang, Y. Li, W. Chen, "Real-time Voxel-Wise Patient-Specific Monte Carlo Dose Reconstruction in Total Body CT Scan Images using Deep Neural Network," 2021.
[9] Shahab S. Band, Atefeh Yarahmadi, Chung-Chian Hsu, "Application of Explainable Artificial Intelligence in Medical Health: A Systematic Review of Interpretability Methods," 2023.