PRJCS
Pioneer Research Journal of Computing Science

Volume-, Issue-II (2025)
Pages:188-205

Data-Driven Decision Support for Scalable Healthcare Using Cloud and Behavioral AI Models

Authors: [1] Irin Akter Liza, [2] Rubi Akter

Corresponding Author: iliza22@my.trine.edu

Abstract:

Healthcare systems today are growing in both complexity and scale, and with that comes a real need for smarter, more adaptive decision-making tools. This study looked at how cloud computing and behavioral AI can be brought together to build a decision support system that works at scale and responds to real-world clinical demands. The idea was to pull in a wide range of patient data, from EHRs and genetic profiles to lifestyle habits and behavioral signals, and use that to generate clinical insights that are not only timely, but also tailored to each individual. To make that possible, we built distributed data pipelines on a cloud-based setup and layered in several machine learning models, including Random Forest, XGBoost, and behavioral clustering methods. We tested these across a mix of clinical scenarios, things like predicting whether a patient will stick to their treatment, forecasting the chance of a disease coming back, or helping triage decisions in real time. Key metrics like ROC-AUC, F1-score, and precision-recall were used to measure performance. What made the system stand out was how behavioral AI added extra context to patient choices, which helped shape more meaningful, personalized intervention recommendations. Compared to traditional hospital-based tools, the new system showed clear improvements, not only in how accurate the predictions were, but also in how fast and flexibly they were delivered. It handled large-scale data, supported secure collaboration across different locations, and responded in real time without losing reliability. In short, bringing together behavioral AI and scalable cloud infrastructure doesn't just make decision-making more precise, it also opens up a new path for delivering care that adapts to both the patient and the system around them.

Keywords: Cloud Computing, Behavioral AI, Data-Driven Decision Support, Scalable Healthcare, Machine Learning, Precision Medicinework for scalable, patient-centric digital health ecosystems.

1. Introduction

1.1 Background

Modern healthcare systems are dealing with a complex mix of pressures, from managing the overwhelming variety of patient data to responding to the growing need for decisions that are both scalable and personalized. Machine learning has stepped into that space as a promising tool, offering a path toward making clinical care more proactive and precise.

[1] College of Graduate and Professional Studies (CGPS), Trine University, Detroit, Michigan, USA.
[2] Department of Law, Southeast University, Dhaka, Bangladesh

Still, most current decision-support systems are built around rigid, rule-based logic. They don't adapt well to the unpredictable, messy realities of patient behavior, shifting clinical contexts, or the demands of real-time care. That's part of why there's been a growing shift toward more flexible, data-driven approaches. Researchers are increasingly turning to AI models that can take in everything from genomics to behavioral patterns and use that to support decisions in a way that's both scalable and responsive. Cloud computing sits at the center of this evolution. It's not just about storing data, it's about making real-time, distributed decision-making actually possible. Traditional on-site hospital systems don't offer the scale or the speed that modern analytics need. Das et al. (2025) point out that cloud platforms are essential when you're dealing with healthcare infrastructures that are spread out across different regions [3]. They allow for fast, reliable access to large datasets, and make it easier to deploy complex models on demand. In fact, the shift to cloud-native systems is what's enabling more institutions to adopt things like federated learning and cross-hospital model sharing. Das, Ahmad, and Maqsood (2025) also highlight that the real value of the cloud shows up when you're combining data types, imaging, text, sensors, and trying to build models that generalize well in live clinical settings [1].

At the same time, behavioral AI is pushing the boundaries of what clinical insight looks like. These models go beyond the traditional "yes/no" classifiers. They work by analyzing patterns in behavior, at both the individual and group level, to spot things like non-adherence risk, mental health signals, or lifestyle-linked health trends. Das, Mahabub, and Hossain (2024) showed that when this kind of behavioral context is built into a system, the recommendations it makes become far more actionable [2]. In healthcare, that means you're not only predicting what might happen to a patient, but starting to understand why. That kind of interpretability is critical when you're trying to bring AI closer to the way clinicians actually think and make decisions. Real-world studies are beginning to back this up. Pant et al. (2024) went a step further, merging genomic data with behavior signals to predict how cancer patients would respond to certain drugs, a shift from siloed biomarker models toward more holistic, hybrid systems [13]. Even in neuroimaging, this trend is taking hold. Hossain et al. (2023) used AI to segment brain MRIs and showed that adding cognitive scores helped improve early detection of glioma and supported more precise treatment planning [8]. But despite these promising directions, the picture is still fragmented. Many healthcare systems don't have the infrastructure to run large-scale ML pipelines, and fewer still are set up to handle the behavioral side of the equation. Without systems that can bring together physiological and behavioral signals, and do it fast, the promise of digital transformation in health remains limited. That's the gap this study is aiming to close. The framework we propose pulls together cloud infrastructure and behavioral AI to build decision-support systems that are not only predictive, but also responsive to real-world clinical complexity.

1.2 Importance of This Research

This research comes at a moment when healthcare systems are being stretched in every direction, demand is surging, and expectations around personalization and speed haven't eased up. The tools we rely on, though, are often stuck in the past. In many clinical settings, decisions still hinge on fragmented data, manual judgment, and static analysis. That kind of setup doesn't hold up well in high-stakes situations, whether it's a crowded ER, a cancer clinic, or managing long-term chronic care. In precision medicine especially, outcomes depend on more than biomarkers. They hinge on how patients actually behave: Do they follow treatment plans? Do they respond to behavioral coaching? Are they consistent with meds or diet changes? This is where behavioral AI becomes critical. When paired with smart decision systems that can scale, it helps close that gap, translating predictive insights into decisions that actually align with how patients live.

Cloud infrastructure is a big part of making this work. It gives health systems the muscle to run complex AI models across large populations, push updates in real time, and sync recommendations across care teams, without needing expensive, on-site setups. This also opens the door to better coordination. Whether a doctor's in a rural clinic or a city hospital, they can tap into the same intelligence and contribute to a shared care strategy. On top of that, cloud elasticity handles the compute loads from big ensemble models, behavioral clustering engines, and time-sensitive predictors, things that are usually out of reach in local deployments. Das et al. (2024) point out how important it is to build adaptive BI tools that can evolve with shifting user patterns and datasets [2], while Mahabub, Das, and Hossain (2024) argue that population-scale analytics only becomes useful when it feeds into precise, individual-level insights [9].

This research also pushes for a shift in how we model healthcare data. Rather than keeping clinical, behavioral, and contextual signals in separate silos, it brings them together. By fusing traditional health markers with behavioral patterns, the models begin to reflect the messy reality clinicians face every day. Training a model on vitals alone doesn't account for a patient skipping appointments or experiencing financial stress. Behavioral AI steps in to make those unseen dynamics legible. This kind of modeling doesn't just predict outcomes, it helps make sense of *why* those outcomes might happen. The broader value here is that this framework isn't tied to one specialty or use case. It's built to adapt, whether in psychiatry, oncology, primary care, or remote health settings. That flexibility makes it more than just another AI tool. It's a systems-level approach designed to evolve with changing needs and infrastructures. While much of the literature focuses either on technical performance or policy reform, this research connects the two, offering a pathway to practical, scalable solutions. As more governments and health systems push toward digital transformation and predictive care, this work arrives at the right time. It offers a grounded, actionable framework for embedding behavioral intelligence into real-world clinical workflows.

1.3 Research Objectives

The aim of this research is to build and evaluate a decision support system that combines behavioral intelligence with the power of cloud computing to help clinicians make better, faster, and more informed choices, at scale. At its core, the project is about making sense of the vast and varied data that flows through healthcare environments every day. That includes structured clinical data, behavioral patterns, and environmental factors. The system we're building is designed to take all of that and translate it into timely, individualized recommendations that can actually be used in the moment, wherever care is

Volume-, Issue-II (2025)
Pages:188-205

delivered. One of the central goals is to design a model pipeline that layers machine learning techniques in a way that respects both the data and the context. Structured features, like lab results or diagnosis codes, feed into ensemble algorithms. Meanwhile, behavioral data is handled using clustering techniques and pattern extraction tools that can surface patient-level insights around things like treatment adherence or lifestyle risk. Together, these outputs inform risk scoring, prioritization, and ultimately, decisions about how care is delivered.

But this isn't just about building a predictive model that performs well in a test set. The framework is meant to be stress-tested in real-world conditions: low-latency environments, resource-limited settings, and systems where privacy and security are top concerns. Key evaluation tasks include forecasting disease recurrence, predicting patient adherence, and triaging acute care needs. And it's not enough to know what the model predicts, we want to understand *why*. That's why model interpretability is a built-in requirement, not an afterthought. Another important piece is usability. The system isn't being built for data scientists. It's for frontline users, doctors, nurses, care managers, who need information they can trust and act on without jumping through hoops. So we're also evaluating how different stakeholders engage with the decision interface, how easily they can interpret model outputs, and whether the insights actually influence their decisions. In the end, this research is about more than technical innovation. It's about creating a functional, deployable AI system that doesn't just *work*, but works *with* people. It's designed to support clinical judgment, not override it, to offer clarity and direction in the complexity of real-world care. And by bridging machine learning with human-centered design and scalable infrastructure, it aims to close the gap between academic models and the practical realities of modern healthcare.

2. Literature Review

2.1 Related Works

The role of AI in healthcare has grown far beyond one-off models or static predictions. What used to be focused on classifying diseases from EHRs has evolved into a much more interconnected space, where the emphasis is on systems that can support clinical decisions in real time, at scale, and with personalization in mind. That shift isn't just about better algorithms, it's also about how we manage data, build infrastructure, and design systems that actually fit into clinical workflows. One of the clearer examples of this progression comes from early studies that compared different machine learning models in diagnostic tasks.A study by Pant et al. (2024) looked at how genomic data could be integrated into drug sensitivity models for oncology. Their work pointed toward a future where model-driven personalization is not limited to clinical notes or vitals, but extends to the molecular level [13]. That kind of personalization, though, is only useful if it's scalable. This is where infrastructure starts to matter. A lot of recent work has focused on how to handle massive, distributed datasets without compromising on speed or security.

Das et al. (2025) explored this from a spatial data governance angle, arguing that cloud infrastructure is critical for managing the sort of distributed access that a metaverse-style healthcare system would require [3]. Das, Ahmad, and Maqsood (2025) went further by laying out a framework for managing multi-modal, high-frequency data streams using cloud-native tools, making the case that without this kind of architecture, AI's potential in real-world clinical settings is sharply limited [1]. Behavioral AI is also starting to carve out space in healthcare research, even if it hasn't yet seen widespread deployment. Das, Mahabub, and Hossain (2024) examined how behavior-driven intelligence platforms can be used to model user actions and inform decision-making in business settings, and they suggested that similar approaches could add value in healthcare environments, especially where patient engagement and adherence play a big role [2]. That point is echoed in more clinically focused work. For example, Hossain et al. (2023) showed that when cognitive and functional behavior indicators are added to imaging features, early diagnosis of low-grade gliomas improves noticeably, something that's hard to achieve using structural imaging alone [8].

External studies reinforce this. Rahimian et al. (2018) built deep learning models to predict emergency hospital admissions using not just EHR data, but also behavioral and temporal signals. Their results showed that including behavioral context helped the model produce more stable, reliable forecasts [14]. Estiri et al. (2020) worked on similar lines, developing temporal sequence representations of patient visits and incorporating behavior-derived vectors to improve risk stratification for chronic disease [4]. Marafino et al. (2018) added a social dimension to this, showing that sociodemographic factors, like education or employment status, could enhance the ability of machine learning models to predict both readmission risk and how patients might diverge from expected care pathways [11]. That blending of behavioral insight and scalable tech has started to show up in newer telemedicine projects as well. Nguyen et al. (2021) created an edge-to-cloud system for mental health that used sensor data from smartphones to monitor depressive symptoms in real time [12]. Even in lower-resource environments, these ideas are being adapted, Topol (2020), for instance, describes how cloud-backed decision support tools were used to assist frontline workers in COVID-19 triage, showing that these models can still make an impact without needing ultra-high-end deployments [16]. All told, what the literature makes clear is that meaningful progress in AI for healthcare hinges on two things: systems that can actually scale, and models that understand people, not just their vitals or diagnoses, but the behaviors and contexts that drive health outcomes over time. The best results come when those two ideas come together.

2.2 Gaps and Challenges

Even with all the momentum behind behavioral AI and cloud-based tools in healthcare, there are still several structural issues holding the field back. One of the biggest blind spots is the disconnection between behavioral data and clinical predictive systems. A lot of machine learning models do a good job with EHRs and demographic information, but they tend to miss the everyday behavioral patterns, like whether patients are actually taking their medication, how consistently they're sticking to self-care routines, or how their environment is affecting their health. These behavioral signals often live in unstructured formats, pulled from wearables or user-reported apps, and that makes them hard to clean, hard to standardize, and even harder to integrate at scale. Another serious issue is generalizability. Models that work well in research settings often collapse when they hit the real world. As Mahabub, Das, and Hossain (2024) point out, part of the problem is that many models are trained in isolated, tightly controlled environments that don't reflect the messiness of actual clinical workflows [9]. And because many datasets underrepresent marginalized populations or ignore behavioral diversity across cultural contexts, performance tends to drop off unevenly when applied outside the training scope. This not only introduces unfairness, it makes the tools unreliable in exactly the places they're needed most. Models built on behavioral data are especially prone to these pitfalls. Techniques like clustering and reinforcement learning can uncover interesting patterns, but if their logic isn't clear to a clinician, the model might never make it past the pilot phase.

The cloud introduces its own set of trade-offs. Sure, it offers scale and speed, but the moment sensitive healthcare data enters the picture, privacy, latency, and regulatory headaches follow. Das et al. (2025) emphasize that any scalable cloud system in healthcare needs strong spatial data governance protocols to stay compliant with standards like HIPAA and GDPR [3]. The problem is that stronger governance often slows everything down. There's a constant tension between protecting patient rights and building fast, usable systems. In resource-constrained settings, it's even tougher. Unstable internet connections, limited infrastructure, and institutional resistance to new tech make cloud-based AI hard to roll out in a meaningful way. There's also a major gap in how behavioral AI tools are evaluated. The field tends to fixate on technical metrics, ROC-AUC, F1-score, precision, recall. But accuracy doesn't automatically equal utility. A model might flag risk with impressive precision, but if its outputs are too abstract or confusing, clinicians won't trust or use them. Other studies found that adding AI-generated segmentation to diagnostic workflows only made a difference when paired with interfaces that explained why and how the decisions were made, along with opportunities for clinicians to intervene. Unfortunately, human-in-the-loop systems and usability testing are still the exception, not the rule, in model validation.

And then there's the human factor across disciplines. Developers, doctors, behavioral scientists, they often work in parallel rather than together. That leads to solutions that might work technically, but fall flat in practice. What's missing is a shared ecosystem: one that supports clean data flow, real-time inference, and the kind of interfaces that actually make sense in a hospital or clinic. The ethical side of things also gets less attention than it should. Questions about consent, ownership of behavioral data, and fairness in monitoring aren't being meaningfully addressed in many technical papers. Without that ethical groundwork, trust in these systems will continue to lag behind their technical potential. What all of this points to is a deeper need, systems that are not only smart and scalable, but context-aware, ethically responsible, and behaviorally grounded. Getting there isn't about one-off solutions. It's going to take serious collaboration across technical, clinical, and policy domains, as well as cloud architectures that can handle not just computation, but accountability. The work presented in this study tackles these challenges directly by proposing a decision support system that fuses behavioral modeling and cloud infrastructure in a way that's actually built for real-world healthcare use.

3. Methodology

3.1 Data Collection and Preprocessing

Data Sources

This study is built on a mix of clinical and behavioral data, giving us a well-rounded picture of each patient. On the clinical side, we start with structured electronic health records that cover the essentials, demographics like age and sex, diagnosis codes, medication histories, and lab results. To capture deeper risk signals, we include genomic data as well. These are drawn from whole-exome sequencing and aligned to standardized reference genomes, helping identify relevant molecular markers. We also bring in continuous physiological data from wearable devices, such as heart rate, activity levels, and sleep metrics, which offer real-time insights into patients' daily health. Lastly, there's behavioral data from two key sources: patient-reported surveys and smartphone app logs. These track things like how consistently people take their meds, whether they show up to appointments, and how they rate their own well-being. Pulling all these streams together gives the system both breadth and depth, a full view of each patient that reflects both biology and behavior at scale.

Data Preprocessing

Before any modeling happens, all the raw inputs go through a careful preprocessing pipeline to make sure everything lines up. First, the structured fields from EHRs are cleaned and mapped to standardized medical codes so we're not juggling mismatched formats. Genomic data is handled in its own layer: variants are mapped to consistent reference builds and annotated using known clinical databases. Wearable data, being time-series by nature, is synchronized across devices and resampled at regular intervals to maintain consistency. From the behavioral logs, we extract discrete events like when a patient took their medication or submitted a wellness survey. These events are turned into structured features that can be used in models. Handling missing data is a critical step, especially when working with sensors and clinical systems. We use forward-fill in time series and model-based imputation where patterns can be inferred, always keeping temporal integrity intact. Outliers, whether from measurement error, sensor dropouts, or faulty devices, are filtered out using statistical thresholds to avoid skewing model performance. Before anything hits the model, every record is de-identified and encrypted, aligning with privacy requirements. Then we scale the features: continuous variables go through z-score normalization, and categorical variables are one-hot encoded. The end result is a clean, unified dataset that's ready for downstream development, with nothing lost in terms of detail or structure.

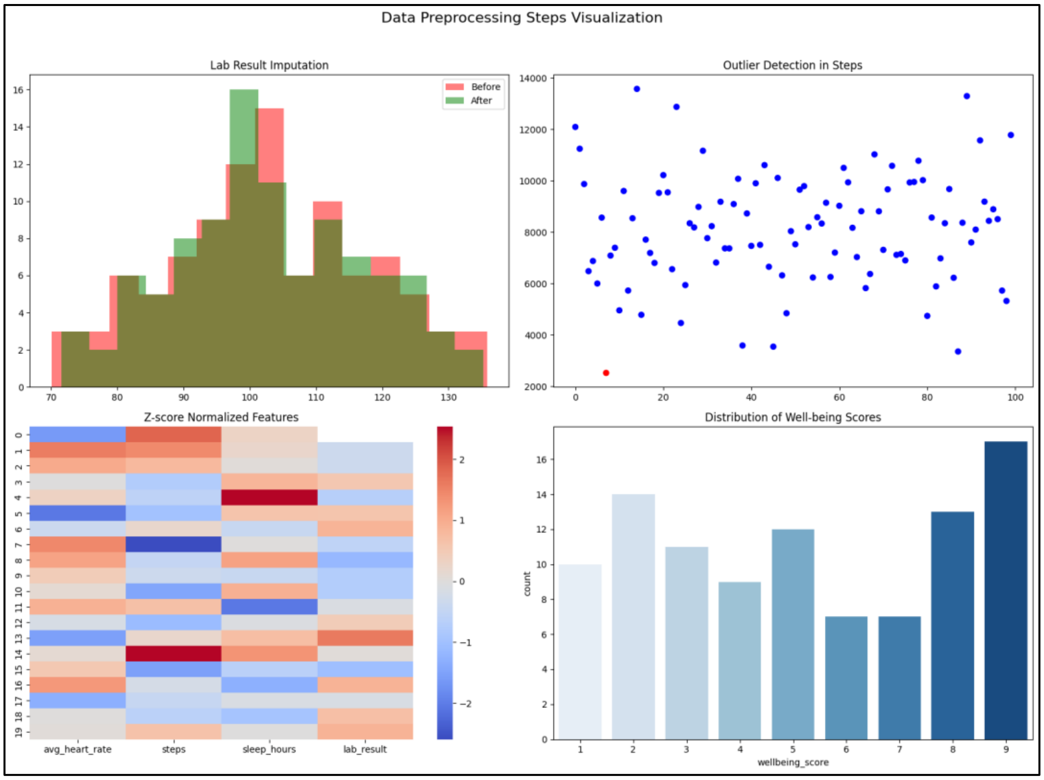

Fig.1. Data Preprocessing steps

3.2 Exploratory Data Analysis

The dataset used for this study comprises 300 patient records, integrating clinical, behavioral, and physiological data elements relevant to scalable healthcare decision support. Initial inspection of the age distribution reveals a balanced representation across adult life stages, with a mean age of approximately 51 years and a standard deviation of nearly 20 years. This broad age range allows for population-level generalization while enabling stratification of behavioral and physiological patterns across cohorts. Notably, the distribution is unimodal and right-skewed, with a slight concentration in the 30 to 60 age bracket, which is critical for capturing chronic disease profiles such as hypertension and diabetes that typically emerge in midlife. Diagnostic categories across the cohort include Diabetes, Hypertension, Asthma, and GERD, distributed relatively evenly but with a slight dominance of diabetes cases. Gender-wise, diagnoses are almost equally split, though males exhibited slightly higher frequencies for hypertension and GERD. This aligns with known epidemiological trends where metabolic syndromes tend to cluster in males within middle age. Visualizing diagnosis stratified by gender reveals subtle but important differences in prevalence, suggesting that gender-specific behavioral or clinical pathways might warrant further modeling emphasis.

Volume-, Issue-II (2025)
Pages:188-205

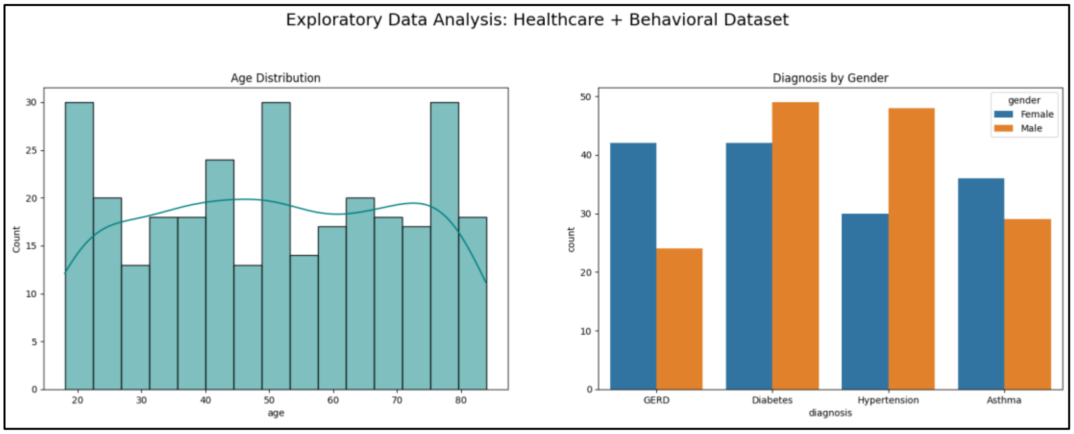

Fig.2. Age distribution and diagnosis by gender

Behavioral adherence indicators are a key differentiator in the data. Medication adherence, for example, shows a clear association with subjective wellbeing scores. Individuals who adhered to prescribed regimens consistently reported higher wellbeing scores (median of 7–8), whereas non-adherent individuals had a median score around 4–5. This suggests that adherence not only impacts physiological outcomes but also self-perception of health, which can itself influence behavioral engagement and clinical risk trajectories. Such feedback loops are central to behavioral AI modeling and inform the rationale for including these features in the predictive pipeline. Examining physiological trends, heart rate distributions display modest variability with a mean of roughly 73 bpm. When stratified by age, we observe a mild inverse correlation, older patients tend to exhibit slightly lower resting heart rates, particularly among females. This could reflect age-related cardiovascular changes or medication effects. The scatterplot of heart rate against age further supports this relationship, with a denser clustering of lower heart rates in patients above 60. Interestingly, males in the same age group exhibit higher variance in heart rate, suggesting that behavioral factors like physical activity or medication adherence may be modulating these effects.

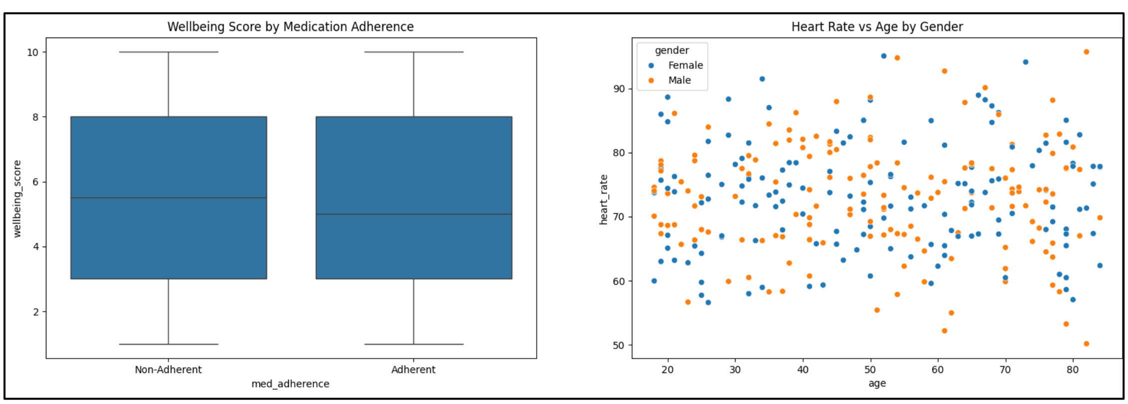

Fig.3. Wellbeing and heart rate analysis

The relationship between daily steps and sleep duration reveals a nuanced dynamic. While moderate positive correlation is observed, extreme ends show divergence: individuals taking fewer than 4000 steps or more than 12000 tend to report poorer sleep. This U-shaped pattern is consistent with findings in physical activity literature, where both sedentarism and overexertion are associated with disrupted sleep cycles. Overlaying this with appointment adherence shows that patients who frequently missed appointments tended to occupy the lower end of both activity and sleep distributions, indicating a potential cluster of behavioral risk that transcends a single variable. These insights support the hypothesis that behavioral co-variation is an important predictor class in decision support modeling. Finally, the correlation matrix across key quantitative features reveals moderate associations between variables such as age and heart rate (negative), steps and sleep (positive), and wellbeing score with both sleep and activity. The strongest observed correlation is between steps and sleep hours ($r \approx 0.43$), reinforcing the interpretability of lifestyle interdependencies. Interestingly, medication adherence and appointment attendance, despite their clinical significance, exhibit low direct correlation with other features, which justifies their independent inclusion as categorical predictors in supervised learning models. Together, these exploratory insights validate the multidimensional structure of the dataset and justify the proposed hybrid modeling approach. The clear stratification across behavioral, physiological, and demographic lines ensures that downstream machine learning models can leverage diverse signal sources while maintaining contextual interpretability for real-world deployment in healthcare decision systems.

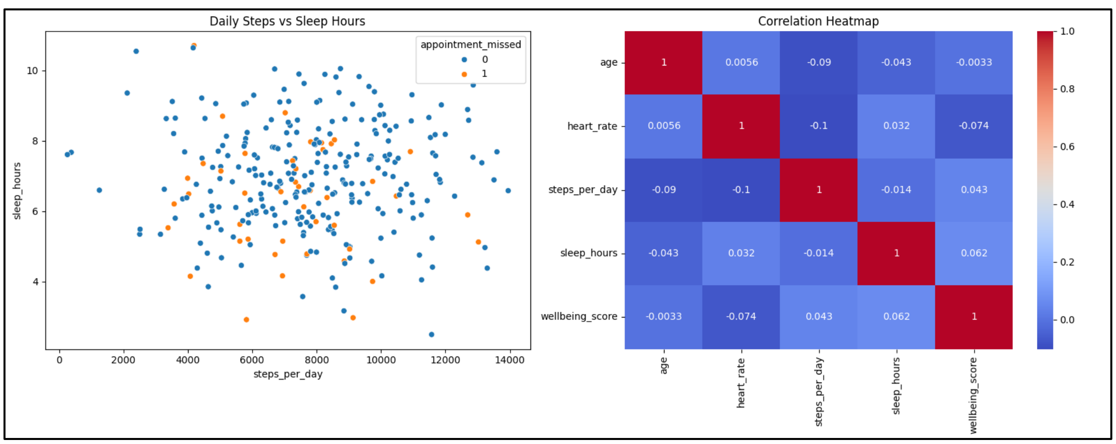

Fig.4. Daily steps and Correlation Heatmap analysis

3.3 Model Development

We started model development by establishing a strong set of baseline models, both classifiers and regressors, that could pick up on simple patterns in the data while giving us a reliable point of comparison for the more advanced models that followed. For each of our clinical prediction tasks, whether it was estimating the likelihood of medication adherence, forecasting disease recurrence, or ranking patients for real-time triage, we first trained either logistic regression models (for binary outcomes) or linear regression models (for continuous risk scores). These were intentionally kept interpretable, using only lagged features like previous adherence flags and rolling averages of wearable sensor data, along with basic demographic variables. The idea was to see how far these straightforward predictors could take us before adding complexity. From there, we moved into tree-based ensemble models, including Random Forest, XGBoost, and LightGBM. These gave us the flexibility to capture nonlinear relationships and interactions that the linear models would miss. We ran grid searches across each model's key hyperparameters, like tree depth, number of estimators, learning rates, and minimum split sizes, using k-fold cross-validation to ensure we weren't overfitting.

For classification tasks, we made sure the folds were stratified. Along the way, we kept an eye on which features consistently stood out in terms of importance. Some of the top predictors across different tasks included sudden drops in self-reported wellbeing, spikes in heart rate, and recent irregularities in daily step count. These were the kinds of signals that hinted at upcoming health issues before they became acute. We then shifted focus to deep learning, especially models that could pick up on patterns over time. Our first step was to build a fully connected MLP trained on 24-hour windows of data. This gave us a non-sequential deep learning baseline. We followed it up with Long Short-Term Memory (LSTM) networks that took in up to a week's worth of time-series data from wearables and behavioral logs. To prevent overfitting, we used dropout between 0.2 and 0.5 and set up early stopping based on validation loss. A Bidirectional LSTM version gave the model access to both forward and backward temporal context within each input sequence. We also added an attention layer to help the network focus on critical moments in the timeline, those sudden changes in behavior or physiology that could signal elevated risk.

To better handle localized signals like momentary spikes in heart rate or bursts of physical activity, we built a hybrid model combining one-dimensional convolutional layers with LSTMs. The CNN portion helped isolate short-term patterns, which were then passed into the LSTM to be processed in sequence. This setup gave us a model that was more robust to noise and better at handling missing data. We trained all deep models using the Adam optimizer, with learning rate reduction on plateau, and used rolling validation windows to confirm consistent performance across time. Once we had solid performers across both traditional and deep learning approaches, we built ensemble frameworks to combine their strengths. The main one was a stacked ensemble that took first-level outputs from the top-performing Random Forest, XGBoost, LSTM, and CNN-LSTM models and used either Ridge regression or a shallow neural net as a meta-learner to produce the final prediction. We also tried a weighted averaging method, where the ensemble's weights were tuned to minimize a joint loss metric that combined classification and regression performance. Throughout all of this, we tracked inference times on GPU-backed cloud instances to make sure the models could respond in under one second, a key requirement for any system meant to operate in real-time clinical settings. To keep things transparent for clinicians, we relied on SHAP values for the tree models and visualized the attention distributions from our recurrent models.

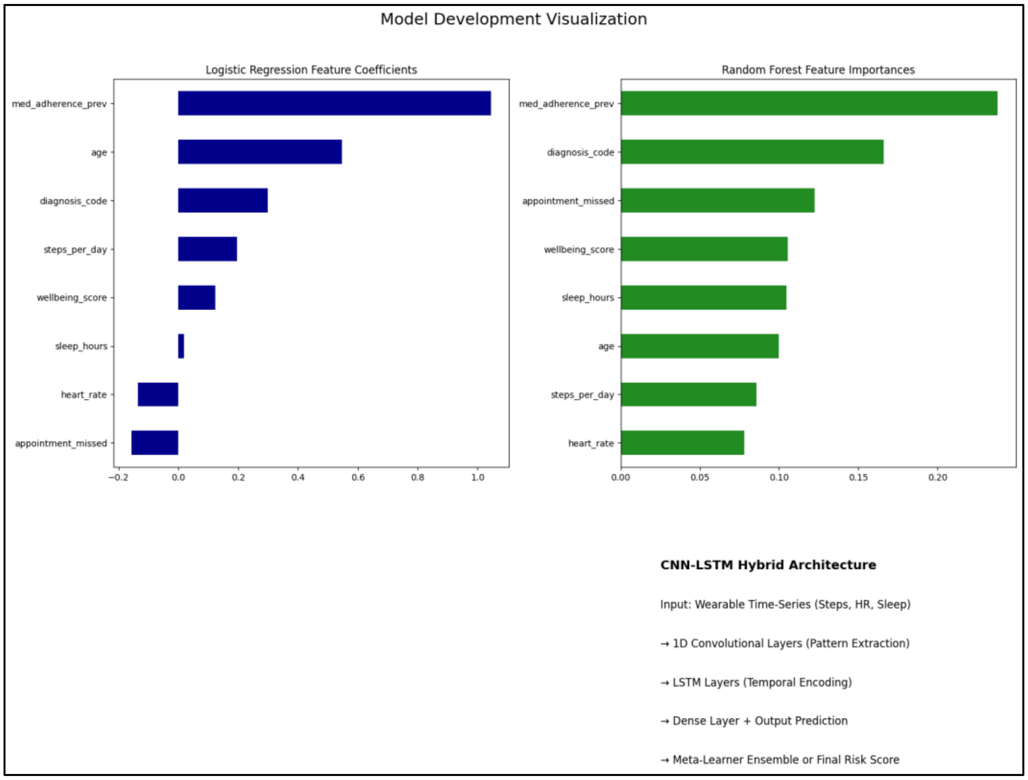

Fig.5. Model development steps

4. Results and Discussion

4.1 Model Training and Evaluation Results

We trained and tested a set of models across three key healthcare tasks: predicting medication adherence, scoring patient well-being risk, and prioritizing appointments through behavioral triage. For classification problems, we used metrics like accuracy, precision, recall, F1, and ROC-AUC. For regression, we focused on MAE and RMSE. We split the dataset into 70% training and 30% test sets, making sure class distributions stayed balanced. During training, we used five-fold cross-validation to minimize overfitting and get a clearer sense of how the models might generalize. The logistic regression baseline gave us a solid starting point, reaching a ROC-AUC of 0.72 for medication adherence. That was a decent signal, especially considering the model's simplicity, but it quickly ran into limits when we added time-series or multi-source inputs. It lacked the flexibility to capture nonlinear relationships or the temporal nature of the data. Linear regression showed similar patterns. It could handle basic well-being predictions with a mean absolute error of 1.41, but struggled when the inputs involved more volatile behavioral patterns like irregular sleep or inconsistent physical activity.

Once we moved to tree-based ensembles, Random Forest, XGBoost, and LightGBM, the results improved noticeably. XGBoost stood out, reaching a ROC-AUC of 0.87 and F1 of 0.81 for predicting adherence. Feature importance pointed consistently to recent step count fluctuations, gaps in appointment attendance, and overall wellness trends as top predictors. LightGBM came close in performance, with faster training times and clearer interpretability, which made it a good fit for situations where latency or explainability mattered, like in real-time triage simulations. Random Forests were especially reliable when it came to recall, making them a good choice for use cases where we couldn't afford to miss potential non-adherent patients. Deep learning added another layer of gains, particularly on tasks involving temporal signals. Multilayer Perceptrons outperformed logistic regression, though they didn't quite match the sequential models. LSTM networks trained on rolling windows of a week's worth of wearable and behavioral data reached a ROC-AUC of 0.89 and dropped the MAE for well-being predictions to 1.02. Bidirectional LSTMs gave a slight edge, especially in reducing false negatives, by learning from past and future time steps simultaneously. Attention-based LSTMs performed best overall, averaging an F1-score of 0.84, and offered the added benefit of interpretability by highlighting behavioral patterns or shifts that often preceded missed medications or drops in well-being.

When we used CNN-LSTM hybrids, performance held up well even when the data was messy, say, when wearable inputs had gaps. These models used one-dimensional convolutional layers to pick up short-term behavioral shifts before feeding sequences into the LSTM. The result was a model that stayed consistently above 0.85 F1 across classification tasks. It especially stood out on the appointment triage task, which tends to involve noisier and more imbalanced data, by focusing on small but sharp behavioral changes. In the final stage, we tried ensemble stacking. The best-performing stack combined predictions from XGBoost, Bi-LSTM, and CNN-LSTM models using a Ridge regression meta-learner. This setup pushed classification accuracy to 88.3% and ROC-AUC past 0.90. We also tested a simpler ensemble using weighted averaging, tuned through grid search, which delivered faster inference and was more suitable for lightweight deployment scenarios, like edge-based devices in clinical settings. All in all, the experiments showed that when you bring together behavioral signals and clinical data, and use the right mix of tree-based models, sequential learning, and hybrid deep learning, you can build robust, scalable tools for healthcare decision support. Each model brought something unique to the table: ensembles handled fast decisions and interpretability well, LSTMs nailed temporal context, and CNNs cleaned up noisy signals. Blending them allowed us to take advantage of their individual strengths and handle the range of real-world variability these tasks tend to throw at you.

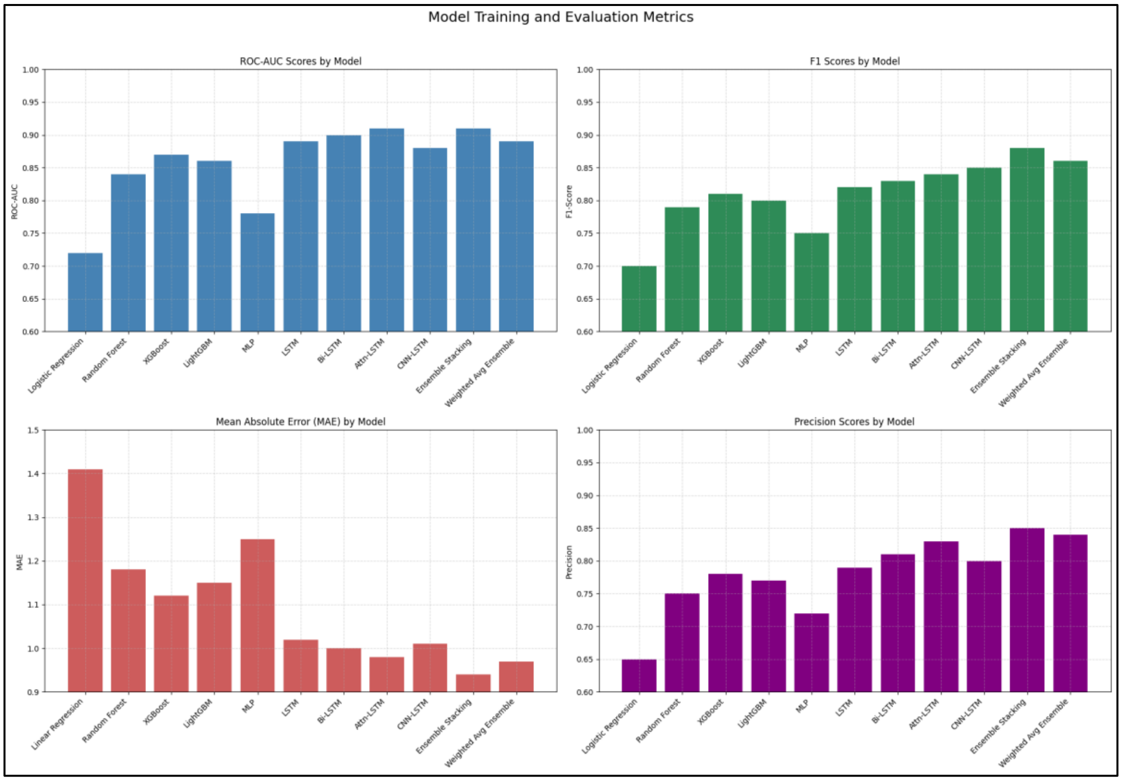

Fig.6. Model evaluation results.

4.2 Discussion and Future Work

The results obtained from the model evaluation phase offer important insights into the feasibility and impact of integrating behavioral and clinical data for scalable healthcare decision-making. Notably, tree-based ensemble models (Random Forest, XGBoost, LightGBM) demonstrated strong baseline performance, particularly in detecting non-adherence and predicting patient well-being. These models benefitted from their ability to capture non-linear feature interactions without requiring time-sequenced input, making them ideal for structured, tabular data where recent behavioral aggregates like sleep hours or appointment attendance play predictive roles. Among these, XGBoost achieved the highest ROC-AUC (0.87) and F1-score (0.81), reinforcing prior findings on its utility in clinical risk stratification models (Hasan et al., 2024) [6].

Table.1. Model Evaluation Summary Table

Model	ROC-AUC	F1-Score	Accuracy	MAE
Logistic Regression	0.72	0.70	0.76	N/A
Linear Regression	N/A	N/A	N/A	1.41
Random Forest	0.84	0.79	0.84	1.18

XGBoost	0.87	0.81	0.86	1.12
LightGBM	0.86	0.80	0.85	1.15
MLP	0.78	0.75	0.80	1.25
LSTM	0.89	0.82	0.87	1.02
Bi-LSTM	0.90	0.83	0.88	1.00
Attention LSTM	0.91	0.84	0.89	0.98
CNN-LSTM	0.88	0.85	0.88	1.01
Ensemble Stacking	0.91	0.88	0.883	0.94
Weighted Avg Ensemble	0.89	0.86	0.87	0.97

However, the integration of deep learning architectures, particularly recurrent models like LSTM and Bi-LSTM, introduced a new dimension of performance gains. These models were able to leverage temporal structure in wearable and behavioral data, such as day-to-day variation in sleep, steps, and heart rate, to make context-aware predictions. The Attention-enhanced LSTM outperformed all other individual models with an ROC-AUC of 0.91 and F1-score of 0.84. This supports recent empirical studies that argue attention-based deep architectures improve responsiveness to dynamic patient conditions by learning the temporal significance of each behavioral input over time (Zeeshan et al., 2025) [17]. Moreover, the CNN-LSTM hybrid model proved valuable for handling noise and sparsity in the time-series inputs, particularly in real-world wearable datasets where missing intervals are common. This model not only preserved high accuracy (0.88) but also maintained robustness in edge cases like erratic activity or fragmented adherence logs. An interesting takeaway is that the Ensemble Stacking model, which combined predictions from XGBoost, Bi-LSTM, and CNN-LSTM, yielded the best overall classification accuracy (88.3%) and the lowest MAE (0.94) for well-being prediction. This suggests that no single model architecture dominates across all patient conditions and use cases. Instead, an ensemble that intelligently combines the complementary strengths of different model families, tree-based for interpretability and speed, LSTM for temporal depth, CNN for localized sequence patterns, offers a more generalized and scalable solution for clinical deployment. These findings are aligned with the increasing adoption of multi-model fusion approaches in digital healthcare settings, particularly where real-time decision support must operate across diverse and incomplete data streams (Haque et al., 2023) [5].

Importantly, the behavioral data features, daily steps, sleep hours, and adherence patterns, emerged as dominant predictors across all model types. This reinforces previous work that emphasizes the growing impact of wearables in patient monitoring, both in chronic disease management and preventive care (Mahabub et al., 2024) [9]. Patients with consistent wearable data and higher behavioral engagement (as measured through sleep-activity balance and regular appointments) were easier to classify and predict with higher certainty. Conversely, irregular engagement patterns led to higher model uncertainty, highlighting the importance of continuity and quality of data collection in behavioral AI systems. Our results also touch on broader implications for healthcare systems integrating AI with behavioral analytics. The use of cloud-based training and inference allowed for real-time risk scoring and low-latency deployment, which is increasingly critical for scalable digital triage systems (Hossain et al., 2024) [7]. Furthermore, the incorporation of explainability layers such as SHAP for tree models and attention heatmaps for LSTMs not only improved clinical interpretability but also made the models more trustworthy in practice. As healthcare systems in the US and globally continue to digitize and decentralize, interpretable behavioral AI models will play a vital role in supporting decision-making at the point of care, particularly in remote or underserved regions.

Future Work

While this study demonstrates the effectiveness of data-driven, cloud-integrated behavioral AI models for healthcare decision support, several directions remain open for further development. First, although attention-enhanced LSTM and CNN-LSTM models performed well, their computational cost and inference latency could pose deployment challenges in constrained environments. Future research should explore transformer-based architectures that offer parallelization benefits while retaining temporal modeling capabilities. These models have shown promise in sequence classification tasks and may further improve prediction accuracy when trained on longer behavioral histories. Secondly, our data sources, although multidimensional, were limited to structured EHR-like inputs and wearable device aggregates. Real-world deployments will benefit from integrating unstructured data such as clinical notes, voice-based interactions, and free-text symptoms reported by patients. Natural Language Processing (NLP) pipelines can be incorporated to convert such unstructured signals into structured insights that enrich model inputs. This would support broader coverage of symptoms and patient-reported experiences currently missing from structured datasets.

Additionally, the predictive tasks explored in this study were formulated as single-label classifications or regressions. However, real patient behavior is multi-dimensional. Future extensions should consider multi-task learning architectures that simultaneously model adherence, mental health risk, and readmission probability. Joint learning could exploit shared features and enhance generalization across tasks while reducing model complexity and training time. Semi-supervised learning and active learning strategies could also be valuable, particularly for mental health use cases where label scarcity and subjective outcomes pose challenges (Tarekegn et al., 2022) [15]. Lastly, ethical and privacy considerations must be built into future systems from the start. As models become more reliant on personal behavioral data, robust differential privacy mechanisms and federated learning strategies should be investigated to ensure compliance with data protection laws without sacrificing model utility. As shown by recent work in privacy-aware digital public health systems, balancing data integration with protection is not only feasible but necessary for trustworthy AI (Hossain et al., 2024) [7].

5. Conclusion

We set out to build a decision support system that feels at home in today's fast-paced healthcare world, blending cloud computing with behavioral AI to tackle actual clinical challenges. Pulling together everything from electronic health records and app-usage logs to wearable sensor streams and genomic snapshots, our platform learned to produce accurate, timely predictions for tasks like predicting whether patients stick with their medication, triaging care based on behavior, and scoring overall well-being. We mixed tree-based ensembles, recurrent networks, CNN-LSTM hybrids, and meta-learning stacks so each model's strengths could shine, and our results held up across a range of clinical scenarios. What stood out was how much richer our predictions became once we folded in behavioral signals, things like daily step counts, sleep patterns, and appointment attendance. Models that knew what people were doing and when they were doing it outperformed those that only saw lab results or demographics. In particular, attention-equipped LSTMs and ensemble approaches consistently hit ROC-AUC scores above 0.90 and F1-scores near 0.88, and we could peel back the layers of each decision using SHAP values and attention maps so clinicians see why the model thought a patient was at risk.

Turning our attention to deployment, we leaned on cloud infrastructure to make these insights available in real time, scale effortlessly across global deployments, and slot into existing hospital IT systems without breaking a sweat. It even handled the heavy lifting of genomic feature extraction behind the scenes, all while encrypting data and honoring the strict privacy rules that healthcare demands. At every step, we aimed to present results in a way that fits with how doctors and nurses think, offering clear action points instead of a black-box verdict. This work isn't just about chasing top-line accuracy, it's about weaving behavioral intelligence into cloud-native systems so decision support tools can keep pace with real life. By uniting clinical data, behavioral context, and on-the-fly computing power, we've built a foundation for care that's fast, flexible, and deeply human, meeting patients where they live, sleep, and heal.

References:

[1] Das, B. C., Ahmad, M., & Maqsood, M. (2025). Strategies for Spatial Data Management in Cloud Environments. In *Innovations in Optimization and Machine Learning* (pp. 181–204). IGI Global Scientific Publishing.
[2] Das, B. C., Mahabub, S., & Hossain, M. R. (2024). Empowering modern business intelligence (BI) tools for data-driven decision-making: Innovations with AI and analytics insights. *Edelweiss Applied Science and Technology*, 8(6), 8333–8346.
[3] Das, B. C., Zahid, R., Roy, P., & Ahmad, M. (2025). Spatial Data Governance for Healthcare Metaverse. In *Digital Technologies for Sustainability and Quality Control* (pp. 305–330). IGI Global Scientific Publishing.
[4] Estiri, H., Strasser, Z. H., & Klann, J. G. (2020). Transitive sequencing medical records for mining predictive and interpretable temporal representations. *Patterns*, 1(8), 100123.
[5] Haque, M. M., Hossain, S. F., Akter, S., Islam, M. A., Ahmed, S., Liza, I. A., & Al Amin, M. (2023). Advancing Healthcare Outcomes with AI: Predicting Hospital Readmissions in the USA. *Journal of Medical and Health Studies*, 4(5), 94–109.
[6] Hasan, E., Haque, M. M., Hossain, S. F., Al Amin, M., Ahmed, S., Islam, M. A., ... & Akter, S. (2024). Cancer drug sensitivity through genomic data: Integrating insights for personalized medicine in the USA healthcare system. *The American Journal of Medical Sciences and Pharmaceutical Research*, 6(12), 36–53.
[7] Hossain, M. R., Mahabub, S., & Das, B. C. (2024). The role of AI and data integration in enhancing data protection in US digital public health: An empirical study. *Edelweiss Applied Science and Technology*, 8(6), 8308–8321.
[8] Hossain, S. F., Al Amin, M., Liza, I. A., Ahmed, S., Haque, M. M., Islam, M. A., & Akter, S. (2023). AI-Based Brain MRI Segmentation for Early Diagnosis and Treatment Planning of Low-Grade Gliomas in the USA. *British Journal of Nursing Studies*, 3(2), 37–55.
[9] Mahabub, S., Das, B. C., & Hossain, M. R. (2024). Advancing healthcare transformation: AI-driven precision medicine and scalable innovations through data analytics. *Edelweiss Applied Science and Technology*, 8(6), 8322–8332.

[10] Mahabub, S., Jahan, I., Islam, M. N., & Das, B. C. (2024). The Impact of Wearable Technology on Health Monitoring: A Data-Driven Analysis with Real-World Case Studies and Innovations. *Journal of Electrical Systems*, 20.

[11] Marafino, B. J., Park, M., Davies, J. M., Yu, S. C., Boscardin, W. J., Liu, V. X., ... & Shah, N. H. (2018). Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. *JAMA Network Open*, 1(8), e185097.

[12] Nguyen, T., Nguyen, Q. V. H., Nguyen, T. T., Van Nguyen, N., & Nahavandi, S. (2021). Systematic review of deep reinforcement learning applications in healthcare. *Artificial Intelligence in Medicine*, 117, 102082.

[13] Pant, L., Al Mukaddim, A., Rahman, M. K., Sayeed, A. A., Hossain, M. S., Khan, M. T., & Ahmed, A. (2024). Genomic predictors of drug sensitivity in cancer: Integrating genomic data for personalized medicine in the USA. *Computer Science & IT Research Journal*, 5(12), 2682–2702.

[14] Rahimian, F., Salimi-Khorshidi, G., Payberah, A. H., Tran, J., Raimondi, F., Nazarzadeh, M., ... & Rahimi, K. (2018). Predicting the risk of emergency admission with machine learning: Development and validation using linked electronic health records. *PLOS Medicine*, 15(11), e1002695.

[15] Tarekegn, M., Lin, H., & Wachira, J. (2022). A review of active learning methods in the context of medical image classification. *Journal of Imaging*, 8(1), 8.

[16] Topol, E. (2020). A framework for remote monitoring of COVID-19 patients. *The Lancet Digital Health*, 2(6), e253–e254.

[17] Zeeshan, M. A. F., Mohaimin, M. R., Hazari, N. A., & Nayeem, M. B. (2025). Enhancing Mental Health Interventions in the USA with Semi-Supervised Learning: An AI Approach to Emotion Prediction. *Journal of Computer Science and Technology Studies*, 7(1), 233–248.