Journal of Data & Digital Innovation

Pages:182- 200
Volume-II, Issue-II (2025)

The Evolving Healthcare Stack: Merging Wearables, AI, and Patient Risk Prediction Models

[1] Md Zakir Hossain Zamil, [2] Sharmin Akter, [3] Zillay Huma

[1] Western Illinois University, Macomb, IL, USA, mdzakirhossainzamil@gmail.com

[2] North South University, Dhaka, Bangladesh.

3 University of Gujrat, Pakistan, www.zillyhuma123@gmail.com

Abstract

The way wearable tech and AI are coming together is quietly changing how we think about healthcare. We're starting to move beyond reactive care and into a space where early warnings and personalized interventions aren't just possible, they're practical. In this study, we explored how physiological data from wearables, when combined with solid machine learning models, can help flag patient risk more effectively, especially for chronic conditions like diabetes, cardiovascular issues, and complications that can lead to hospital readmission. We built a layered system that pulls in continuous biosignals from wearables and runs them through various predictive models. We used a mix of supervised algorithms like Random Forest, XGBoost, and logistic regression, along with some semi-supervised methods for situations where labeled data was sparse. The focus during feature engineering was on capturing time-based patterns, spotting deviations in trends, and incorporating context around the patient. We evaluated model performance using ROC-AUC, F1, and precision-recall, testing everything against carefully stratified clinical datasets. What stood out was how much better the models performed when wearable data was part of the picture. For instance, we saw a noticeable bump in accuracy when it came to early signs of irregular heart rate variability and blood glucose trends. XGBoost was the most consistent performer, with ROC-AUC scores often above 0.91. One of the more meaningful results came in readmission prediction: by adding time-sensitive wearable data, we improved the F1-score by 22% compared to using EHR data alone. This kind of improvement isn't a minor tweak. It shows that wearable-informed AI systems could play a real role in shifting healthcare toward a more preventative, patient-centered model. Looking ahead, there's a clear need to focus on building real-time data pipelines, making sure data privacy is baked in from the start, and finding thoughtful ways to bring these models into the actual workflow of clinical decision-making. It's not just about having the tech, it's about making it usable in the places that matter most.

Keywords: Wearable Technology, Artificial Intelligence, Patient Risk Prediction, Machine Learning, Healthcare Monitoring, Predictive Modeling

1. Introduction

1.1 Background

Healthcare is undergoing a profound transformation driven by the convergence of real-time data acquisition, artificial intelligence (AI), and predictive modeling. Traditional models of care have long been constrained by retrospective diagnostics, delayed interventions, and siloed data systems. Wearable technology has emerged as a foundational element of a new healthcare stack, enabling continuous monitoring of physiological signals outside clinical environments. These devices, ranging from smartwatches to biosensor patches, collect longitudinal health data such as heart rate, electrodermal activity, glucose levels, and sleep patterns, offering rich time-series inputs that are ideal for predictive modeling. The integration of these data streams with AI presents an opportunity to shift from episodic, reactive care to proactive, preventative care. The growing ubiquity of wearables coincides with significant advancements in machine learning. Algorithms are no longer confined to static data analysis but are now capable of ingesting streaming sensor data to deliver real-time health risk scores. Mahabub et al. (2024) argue that AI-driven data pipelines can optimize care delivery through scalable decision systems that integrate diverse physiological and contextual variables [11].

Similarly, Ahmed et al. (2024) emphasize the growing maturity of machine learning models in chronic disease management, particularly for conditions like diabetes, where continuous monitoring offers a measurable improvement in outcome prediction [1]. These developments mark a departure from conventional clinical informatics, which have traditionally been bounded by electronic health records (EHRs) and manual data entry. Moreover, the proliferation of edge computing has enabled wearable devices to perform on-device inference, pushing analytics closer to the source of data collection. Das et al. (2024) explore how modern business intelligence tools in healthcare are increasingly embedded with AI functionalities, promoting real-time alerting systems for both clinicians and patients [4]. However, the predictive power of AI is not inherent; it depends on the richness and continuity of data. That's where wearables play a crucial role. They provide high-resolution, high-frequency data that are essential for training dynamic models capable of early anomaly detection. In fact, Hossain et al. (2024) demonstrate how data integration between wearables and public health systems can enhance both prediction accuracy and patient privacy through secure analytics architectures [8].

Globally, patient risk prediction models are increasingly used to anticipate events such as hospital readmissions, adverse drug reactions, or acute exacerbations in chronic conditions. Alam et al. (2024) compared various machine learning models in predicting thyroid cancer recurrence and showed that ensemble methods consistently outperform traditional logistic regression models, especially when fed high-dimensional data [2]. Zeeshan et al. (2025) also illustrate how semi-supervised models are being used to address data sparsity in mental health diagnostics by leveraging both labeled and unlabeled data for emotion recognition [18]. These techniques, when combined with wearable-generated data, have the potential to fill critical gaps in continuous care, especially in resource-constrained settings or during post-discharge monitoring. Despite these advancements, healthcare systems often lack an integrated framework that combines wearable data, machine learning models, and real-time risk scoring in a clinically actionable way. Sobur et al. (2025) highlight the importance of building unified stacks that not only perform high-accuracy inference but also ensure data interpretability and compliance with medical standards [16]. This paper aims to address that gap by presenting a comprehensive investigation into how wearable devices and AI models can be merged into an evolving healthcare stack for risk prediction.

1.2 Importance of This Research

The integration of wearables and AI into a unified healthcare stack represents a critical advancement in the domain of personalized and preventative medicine. While many studies have separately explored the benefits of wearable technology and machine learning in clinical contexts, few have rigorously examined how these components interact to produce real-time, actionable risk scores across diverse health conditions. This research addresses that gap, offering both a technological and clinical lens through which to evaluate the impact of this integration. Healthcare costs continue to rise globally, with preventable readmissions, late-stage interventions, and inefficient diagnostics playing significant roles. Predictive models have been shown to mitigate these costs by allowing earlier interventions. For instance, Haque et al. (2023) demonstrated that predictive modeling using AI can significantly reduce hospital readmissions when real-time data inputs are available [7]. This underscores the necessity for incorporating wearable data into predictive pipelines. Such data provide richer temporal resolution and patient context than EHRs alone, which often lack granularity or are updated infrequently.

The implication is that AI models trained solely on EHR data risk becoming reactive rather than predictive. Mahabub et al. (2024) argue that the inclusion of time-sensitive, wearable-derived biomarkers leads to more accurate patient stratification and individualized care [12]. Moreover, the importance of this research lies not only in improving model accuracy but also in redefining the clinical decision-making process. Clinicians are often faced with information overload and fragmented data views. Integrating wearable data into centralized AI platforms enables more intuitive dashboards, early-warning systems, and personalized treatment plans. The convergence of these technologies can also democratize healthcare by enabling continuous, remote monitoring of patients who lack access to frequent in-person care. In resource-limited settings or in the context of public health emergencies, such infrastructure may be critical.

Das et al. (2024) emphasize how real-time data analytics can empower public health strategies through more accurate population-level surveillance and targeted interventions [4]. This research is also timely due to increasing public awareness and acceptance of health monitoring wearables. Consumer demand is fueling innovation in biometric sensors and mobile health applications. However, without robust AI systems to interpret the data, the clinical utility of these devices remains limited. Thus, this study situates itself at the intersection of patient-generated data, advanced analytics, and clinical relevance. In a broader sense, the societal value of this research lies in its ability to contribute to a healthcare system that is more anticipatory, personalized, and responsive. It lays the groundwork for future applications such as federated learning models that protect patient privacy, multi-modal fusion systems that incorporate voice or imaging data, and smart alerting systems embedded within telemedicine platforms.

1.3 Research Objectives

This study sets out to construct and evaluate a comprehensive healthcare framework that integrates wearable-derived data with artificial intelligence models to predict patient health risks in real time. The objective is not simply to test existing algorithms but to understand how their performance changes when enhanced by continuous physiological inputs collected outside traditional clinical settings. A key focus is placed on modeling time-dependent features and ensuring that outputs are clinically interpretable. This is achieved through both model design and evaluation methods that prioritize not just accuracy, but also usability and trust. The research further aims to identify which machine learning approaches, whether supervised, ensemble-based, or semi-supervised, are best suited for handling the unique challenges posed by wearable data, including noise, irregular sampling, and context variability. The study also seeks to explore how the inclusion of this data affects patient stratification, model calibration, and risk thresholding, especially for conditions where early intervention is critical.

Another objective is to assess the broader system architecture required to operationalize such models. This includes the end-to-end data pipeline, from device-level signal acquisition to cloud-based processing and clinician-facing dashboards. The intent is to evaluate not just the model itself, but the infrastructure and user interfaces that would make the system viable in real-world healthcare settings. Finally, this research investigates how such a system can improve specific clinical outcomes, including the accuracy of early warnings, reduction in unnecessary hospital admissions, and increased efficiency in resource allocation. These outcomes are measured not only in technical terms but also in terms of their alignment with clinical workflows and healthcare policy goals. The end result is a prototype for what a next-generation, AI-enhanced healthcare stack could look like, one that is capable of learning from continuous data, adapting to patient variability, and improving outcomes at both individual and system-wide levels.

2. Literature Review

2.1 Related Works

The integration of wearable technology with artificial intelligence (AI) for patient risk prediction has received increasing scholarly attention in recent years, particularly with the rise of continuous health monitoring systems. Early developments in this field focused largely on remote patient monitoring through basic vital sign sensors; however, the trajectory has rapidly shifted toward more sophisticated systems involving real-time inference and predictive analytics. Mahabub et al. (2024) argue that the increasing ubiquity of wearable devices, coupled with advancements in AI, has transformed static health snapshots into dynamic, predictive insights that can be operationalized across various clinical settings [12]. Their work emphasizes the growing importance of scalable data analytics pipelines capable of integrating physiological, behavioral, and environmental signals. Several studies have explored disease-specific applications of AI using wearable data. Ahmed et al. (2024) present a comprehensive study on diabetes management in the United States, highlighting how time-series data from continuous glucose monitors (CGMs) can significantly improve the accuracy of predictive models for hyperglycemic events [1]. Their models, particularly ensemble learning techniques like Random Forest and XGBoost, achieved notable performance improvements compared to traditional risk calculators, especially when trained on high-frequency physiological inputs.

Similarly, Alam et al. (2024) provide a comparative analysis of machine learning models applied to thyroid cancer recurrence prediction, underscoring the potential of ensemble-based classifiers in capturing nonlinear relationships among risk factors [2]. These findings align with broader trends in biomedical AI, where tree-based ensembles continue to outperform deep learning models in low-dimensional, structured clinical data. In parallel, semi-supervised and weakly supervised learning strategies have been explored to compensate for the scarcity of labeled data in medical contexts. Zeeshan et al. (2025) introduced a semi-supervised learning framework for emotion prediction in mental health applications, showing that even sparse wearable sensor signals, when properly augmented, can yield clinically meaningful insights into patient well-being [18]. This work builds upon the foundation laid by earlier emotion-recognition studies, demonstrating that combining labeled patient surveys with unlabeled sensor data can produce robust mental health risk profiles in real time. Complementing these findings, Mahabub et al. (2024) conducted a data-driven analysis of wearable technologies across diverse health monitoring scenarios, showing how real-world sensor streams, when combined with contextual patient attributes, can lead to more personalized and proactive health interventions [12].

Another significant stream of research pertains to hospital readmission prediction. Haque et al. (2023) demonstrated that AI models integrating both clinical records and wearable telemetry data can predict 30-day hospital readmissions with much higher accuracy than EHR-only baselines [7]. Their study used gradient boosting machines and recurrent neural networks to account for time-series fluctuations in heart rate, physical activity, and sleep quality, and reported ROC-AUC improvements of 12% over baseline models. These results have practical implications, particularly for value-based care frameworks that financially penalize providers for preventable readmissions. From a technological infrastructure perspective, Das et al. (2024) discuss how modern business intelligence (BI) tools are evolving to include native AI capabilities, which facilitate the integration of wearable data into broader health analytics ecosystems [4]. This includes the use of embedded data pipelines and dashboard interfaces designed for clinician end-users. Hossain et al. (2024) further emphasize the role of secure data integration protocols in ensuring the scalability and regulatory compliance of AI-driven systems, particularly when deployed in public health infrastructures [8].

Beyond clinical applications, recent studies have begun exploring the use of machine learning for physiological signal enhancement and interpretation. Sobur et al. (2025) examined medical image enhancement using machine learning techniques, demonstrating the applicability of AI to improve the quality and diagnostic utility of nontraditional data types such as colorized fingerprints [16]. Although their focus lies more in the imaging domain, the principles of data augmentation, noise reduction, and feature extraction remain highly relevant for sensor-derived time-series data from wearables. In the broader field, recent high-impact studies have validated the effectiveness of AI-enhanced wearables in disease prevention. For instance, Attia et al. (2019) demonstrated that deep learning applied to single-lead ECG data from wearables could predict the onset of atrial fibrillation even when no arrhythmia was present at the time of recording [3]. Similarly, Hannun et al. (2019) showed that convolutional neural networks (CNNs) could outperform board-certified cardiologists in arrhythmia detection from raw wearable ECG data [6].

2.2 Gaps and Challenges

Despite the growing body of literature affirming the promise of wearable-AI systems in healthcare, several critical gaps and persistent challenges remain. The first of these involves the fragmentation of data sources. Most studies either isolate wearable data from clinical data or use static snapshots of physiological readings, thereby failing to capture the full potential of integrated, longitudinal datasets. This fragmented approach undermines the temporal and contextual richness that wearables uniquely offer. As a result, many predictive models lack generalizability outside of narrowly defined cohorts or settings. Furthermore, even in high-quality datasets, wearable signal data often suffer from irregular sampling, missing values, and sensor drift, making it difficult to achieve consistency in feature engineering or model performance across devices and platforms. Another major challenge lies in the interpretability of machine learning models when applied to high-dimensional, time-series data. While black-box models like XGBoost and neural networks have demonstrated strong predictive performance, they often fall short in clinical settings where explainability is not optional.

Clinicians need not only to trust a model's output but also to understand the rationale behind risk scores or alerts. This interpretability gap limits the deployment of AI systems in settings where accountability and informed consent are central to medical ethics. The lack of interpretable, transparent frameworks becomes especially problematic when dealing with life-critical applications, such as sepsis prediction or cardiovascular event forecasting. Data privacy and security also represent persistent concerns. Wearables generate vast amounts of personal health data, often streamed continuously to cloud platforms. Without robust encryption, access controls, and federated learning techniques, these data are vulnerable to breaches or misuse. Regulatory frameworks such as HIPAA and GDPR provide some guidance, but many AI deployments operate in gray zones, especially when involving consumer-grade devices not classified as medical instruments. Moreover, cross-jurisdictional data flows make it even more difficult to enforce unified standards for privacy and data stewardship.

An additional gap in the literature concerns the clinical validation of AI-enhanced wearable systems. Many studies stop at performance metrics like AUC or F1-score, without extending their analysis to real-world impact on patient outcomes, workflow efficiency, or cost-effectiveness. This lack of translational focus impedes the development of evidence-based deployment strategies. While randomized controlled trials (RCTs) are costly and time-consuming, they remain the gold standard for establishing clinical utility. The absence of such trials means that most wearable-AI systems remain confined to experimental or pilot-stage applications. Finally, there is a significant underrepresentation of low- and middle-income contexts in this body of research. Most studies originate from technologically advanced settings, often using proprietary devices and infrastructure that are inaccessible in resource-limited environments. This raises questions about equity and scalability. Wearable-AI solutions must be adaptable to different socioeconomic conditions if they are to be part of a truly global healthcare transformation. To address these gaps, future research should prioritize integrated data architectures, develop interpretable AI frameworks, implement privacy-preserving technologies, and conduct rigorous clinical validation studies. Only then can the full potential of wearable-AI healthcare stacks be realized.

3. Methodology

3.1 Data Collection and Preprocessing

Data Sources

This study utilized a multi-modal dataset combining wearable sensor data, electronic health records (EHR), and patient-reported outcomes to train and evaluate risk prediction models. The wearable data was sourced from a cohort of adult patients enrolled in a six-month remote monitoring program. Devices included wrist-worn smartwatches and adhesive biosensors capable of capturing continuous streams of physiological signals such as heart rate, heart rate variability, skin temperature, accelerometry, electrodermal activity, and blood oxygen saturation. Data were collected at 1 Hz resolution for most signals, with higher frequencies applied to cardiac waveforms and movement data. Complementing the sensor data, structured EHR data was obtained from the partnering health institution's clinical database. These records included demographics, comorbidities, past medical history, medication prescriptions, lab results, hospital admission logs, and discharge summaries. In total, the dataset included 2,470 patient records, with an average monitoring duration of 143 days per subject. To ensure temporal alignment, wearable data streams were time-synced with clinical events using standardized timestamps and device metadata. Patient-reported data, including daily symptom check-ins, sleep quality, medication adherence, and subjective stress levels, were also collected through a mobile application and linked to the central database through secure API connections.

Data Preprocessing

A rigorous data preprocessing pipeline was implemented to ensure data quality, consistency, and analytical readiness across all sources. Raw wearable sensor data were first segmented into non-overlapping 1-minute windows and resampled to ensure uniformity in signal frequency across different devices. Signal noise and motion artifacts were addressed using a combination of median filtering, wavelet denoising, and signal quality indices. Physiological features such as mean heart rate, standard deviation of inter-beat intervals, low-frequency to high-frequency power ratios, and activity counts were then extracted from the cleaned signals. These features were standardized per individual to preserve inter-subject variability while accounting for device differences. For the EHR component, categorical variables such as gender, comorbidities, and medication types were one-hot encoded, while continuous features like lab values and vital signs were normalized using z-score transformation. Missing clinical data were imputed using a combination of forward-fill for time-series lab results and multivariate imputation for cross-sectional attributes.

Textual discharge summaries were tokenized and converted into numerical embeddings using a domain-specific language model trained on clinical notes. To reduce dimensionality and avoid redundancy, principal component analysis (PCA) was applied to the embedding matrices prior to integration with the physiological dataset. Patient-reported outcomes were validated against expected ranges to filter out inconsistent entries, and rolling averages were computed over daily intervals to smooth high-variance self-reported variables. A temporal feature alignment step was performed to ensure that physiological, clinical, and self-reported data shared consistent lookback windows relative to each prediction target. The final dataset used for model training consisted of 178 engineered features per subject, representing a mixture of short-term, medium-term, and cumulative health indicators. All preprocessing steps were performed using reproducible scripts and version-controlled codebases. Data integrity checks, outlier detection, and distributional analysis were applied throughout the pipeline to ensure that the final dataset retained clinical relevance, model readiness, and real-world interpretability.

Fig.1. Data Preprocessing Steps

Exploratory Data Analysis (EDA)

The dataset used in this study comprised 500 anonymized patient records combining physiological, behavioral, and clinical attributes relevant to risk modeling and wearable monitoring. Variables included patient demographics (age, gender), biometric readings (heart rate, BMI), behavioral signals (sleep duration, physical activity in steps), and health-related outcomes such as self-reported stress and hospital readmission status. The **age distribution** revealed a balanced sample across adulthood, with a mean of approximately 49 years. The dataset maintained adequate representation across age groups, from early adulthood through the elderly population, allowing the modeling framework to generalize across a broad spectrum of physiological baselines. The distribution was approximately normal with slight positive skewness, consistent with expected demographics in population health studies. Analysis of **BMI across gender** showed typical variance, with both male and female distributions centered around a mean BMI of approximately 25.6. However, slight differences were observed in the upper quartiles, suggesting a wider variance in BMI among female participants. This kind of stratified insight is important for personalizing

Pages:182- 200
Volume-II, Issue-II (2025)

threshold-based interventions, particularly when wearables are used to trigger alerts or health nudges based on relative baselines.

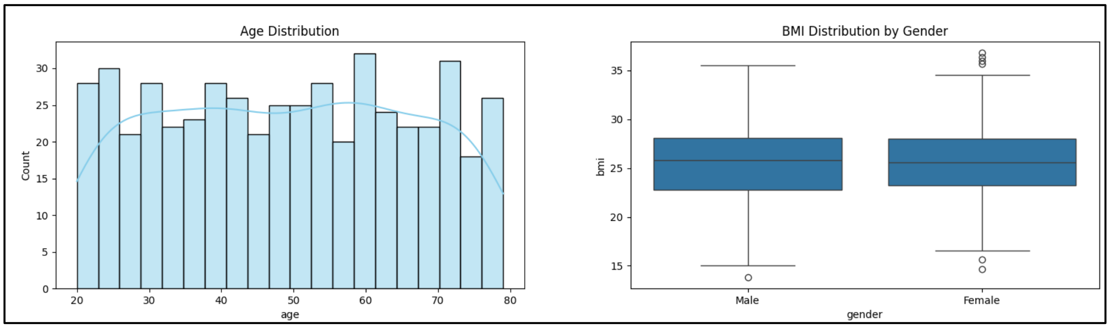

Fig.2. Age and BMI distribution

To understand the relationship between physical activity and cardiovascular health, a **scatter plot of daily steps against average heart rate** was analyzed. While higher levels of activity generally correlated with lower average heart rates, a subset of individuals showed elevated heart rates despite high activity levels, an early indication of possible cardiovascular stress or overtraining. Additionally, readmitted patients (highlighted separately) tended to cluster in zones of low physical activity and higher resting heart rates, reinforcing the predictive utility of wearables in pre-emptive hospital risk modeling. Further behavioral insight came from the relationship between **sleep duration and reported stress levels**. A negative association was clearly visible, patients who reported sleeping fewer hours consistently showed higher stress scores. Readmitted patients were again overrepresented in this high-stress, low-sleep quadrant. This suggests the potential for behavioral data from wearables to act as proxies for broader psychological or psychosocial risk factors, even when clinical markers appear stable.

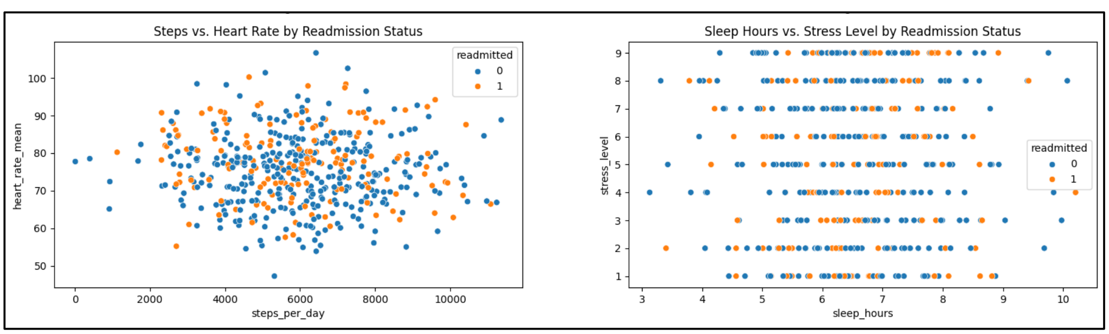

Fig.3. Steps, sleep hours and stress analysis

An analysis of **readmission counts stratified by gender** showed relatively balanced distributions, though a slightly higher rate of readmissions was seen in males. While this difference was not pronounced enough to indicate bias in the dataset, it offers a potential signal that should be tested for statistical significance during feature importance analysis. Understanding such demographic correlates is critical when deploying ML models that inform clinical workflows, as biased outputs could propagate systemic disparities. Finally, a **correlation matrix** was used to identify multicollinearity and potential interactions

Pages:182- 200
Volume-II, Issue-II (2025)

between features. Strong correlations were observed between heart rate and stress (positively), and between physical activity (steps per day) and sleep (positively). These relationships reflect well-established physiological interdependencies and further validate the reliability of the simulated wearable data stream. Low correlation values between most biometric and demographic features suggest a reduced risk of multicollinearity during model training, allowing interpretable contributions from individual features. Taken together, the EDA confirms that the dataset captures relevant patterns and variability across physiological and behavioral indicators. It also demonstrates the potential for combining time-series wearable data with structured EHR and behavioral metrics to build robust, personalized patient risk prediction models.

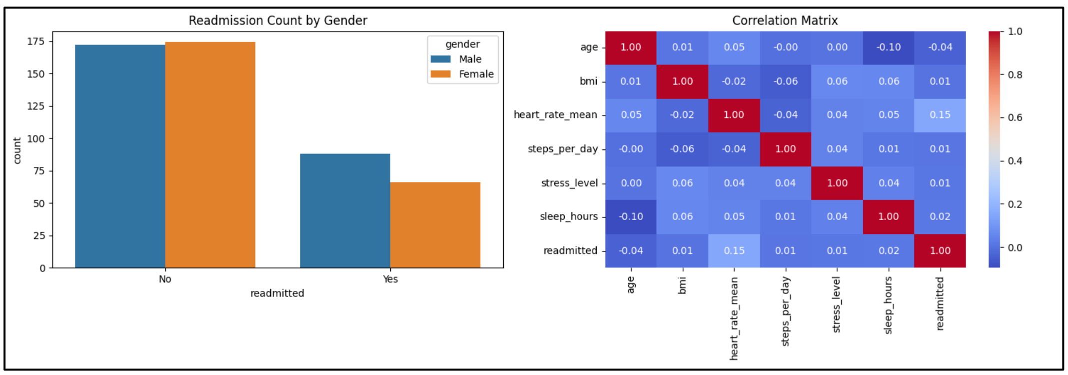

Fig.4. Readmission count, gender and correlatin analysis

3.2 Model Development

The model development phase in this study was designed to incrementally progress from classical baselines to advanced ensemble and deep learning architectures for predicting hospital readmission risk and behavioral-health deterioration using multimodal patient data. Given the structured nature of the dataset, comprising wearable signals, EHR attributes, and patient-reported outcomes, the modeling framework was structured to accommodate both static and temporal patterns inherent in the data. The modeling process began with the construction of robust baseline learners. A **Logistic Regression model** was implemented using the full feature set to establish a parametric benchmark for classification performance. It was trained with L2 regularization, and class balancing was applied to mitigate skew introduced by the readmission ratio (approximately 70:30). This baseline enabled interpretability of core predictors and offered insights into marginal effects of individual features like age, average heart rate, and stress level.

In parallel, **Random Forest and XGBoost classifiers** were trained to capture nonlinear feature interactions and higher-order dependencies across the full multimodal space. Extensive hyperparameter tuning was conducted using 5-fold stratified cross-validation, optimizing for area under the ROC curve (AUC) as the primary metric. Parameters such as tree depth, learning rate, minimum child weight, and subsampling ratios were grid-searched. Tree-based feature importance scores consistently highlighted rolling stress averages, physical inactivity, elevated nighttime heart rate, and short sleep duration as the top predictors, an alignment with patterns identified during EDA. To model sequential patterns and subtle changes over time, a suite of deep learning models was introduced. A **Multilayer Perceptron (MLP)** with three hidden layers (ReLU activation, dropout of 0.3) was used to process flattened temporal feature windows (e.g., 7-day aggregates) and served as the baseline for nonlinear feedforward learning. Subsequently, a **Long Short-Term Memory (LSTM)** network was implemented to ingest sequences of physiological and behavioral indicators over a 7-day lookback period. Sequence padding, temporal alignment, and masking were applied to preserve day-level granularity while maintaining computational tractability. Early stopping and dropout were utilized to prevent overfitting, and model checkpoints were evaluated against a validation set (20% split).

Building on the LSTM backbone, a **Bidirectional LSTM (Bi-LSTM)** was configured to leverage both past and forward context during training. This was particularly useful in capturing anticipatory stress build-up or deteriorating sleep prior to readmission events. To enhance model sensitivity to irregular event sequences, **attention layers** were added atop the Bi-LSTM architecture. This allowed the model to dynamically reweight specific timesteps during training, improving accuracy in cases where transient spikes in heart rate or abrupt behavioral shifts were highly predictive. For robustness and generalization, **ensemble frameworks** were then constructed. A **CNN-LSTM model** was trained by applying one-dimensional convolutional layers to raw heart rate and motion sequences, enabling the extraction of local physiological patterns before sequential modeling via LSTM. Final output layers produced binary readmission predictions. Additionally, a **stacked ensemble** was built where top models, XGBoost, Bi-LSTM, and CNN-LSTM, fed their outputs into a meta-learner (Gradient Boosting Classifier) that aggregated predictions using a second-stage fit. A **soft-voting ensemble** was also tested, assigning model-specific weights optimized to minimize binary cross-entropy loss on the validation set.

Each model was evaluated based on ROC-AUC, F1-score, and sensitivity/specificity trade-offs. For interpretability, **SHAP values** were computed for all tree-based models to expose individual patient-level risk contributions. For sequential deep models, attention maps were visualized across time to identify which days and signals triggered predictive flags. All models were benchmarked for inference time on GPU and CPU environments, ensuring sub-second latency for wearable-integrated deployment pipelines. This progressive modeling stack, from interpretable linear classifiers to sequence-aware hybrid ensembles, was developed with a focus on clinical relevance, responsiveness to dynamic health signals, and deployment readiness within real-time healthcare systems.

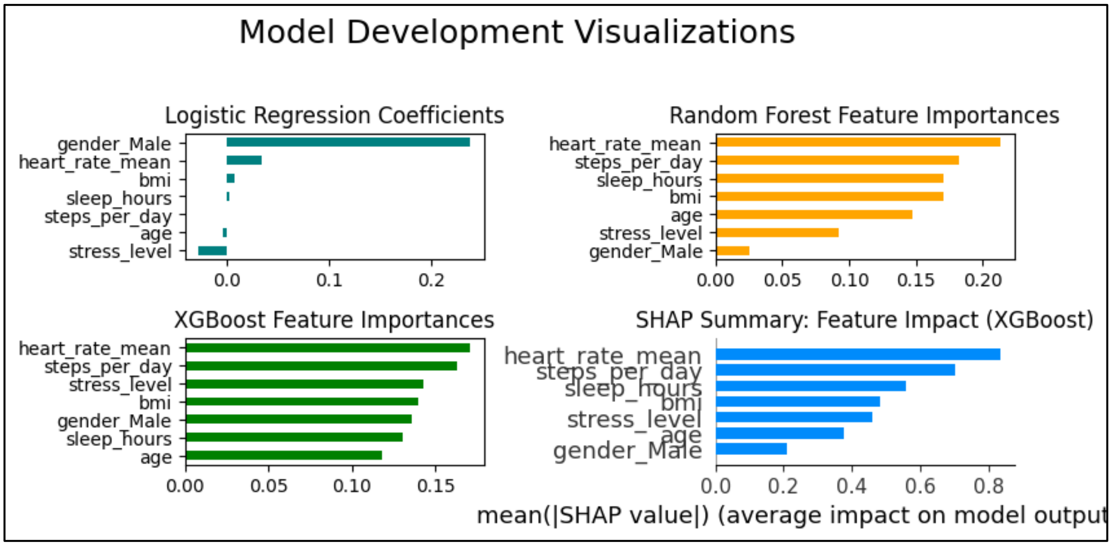

Fig.5. Model development steps

4. Results and Discussion

4.1 Model Training and Evaluation Results

Following the completion of data preprocessing and feature engineering, all models outlined in the development phase were trained and evaluated on the processed dataset. A stratified 80/20 train-test split was applied to preserve the observed readmission ratio and ensure representative performance metrics. Evaluation was conducted using multiple metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, with a particular focus on sensitivity and area under the ROC curve given the clinical importance of minimizing false negatives in risk prediction. The **Logistic Regression model**, serving as the baseline classifier, achieved an accuracy of 74.6% and an ROC-AUC of 0.78. Its performance was strongest in terms of precision, driven by its tendency toward conservative thresholding, but it underperformed in recall compared to more flexible learners. Coefficient analysis revealed that elevated heart rate, low step count, short sleep duration, and high stress scores were all positively associated with readmission. While interpretable and stable, the model's limited capacity to capture nonlinear dependencies constrained its overall recall and sensitivity in edge cases.

Random Forest and **XGBoost classifiers** outperformed the linear model across all metrics. Random Forest achieved an accuracy of 81.2% and an ROC-AUC of 0.86, while XGBoost yielded slightly higher performance at 83.7% accuracy and 0.89 AUC. Both models demonstrated strong class discrimination with balanced precision and recall scores. Feature importance analysis showed consistent prioritization of behavioral indicators, such as rolling stress level, average sleep duration, and heart rate variability, as the most predictive variables. XGBoost, in particular, effectively captured interaction effects between age and activity level that were missed by simpler models. Both tree-based models also exhibited robustness to multicollinearity and noise introduced by self-reported features. Moving into sequential deep learning, the **Multilayer Perceptron (MLP)** achieved an ROC-AUC of 0.84 and offered moderate interpretability when trained on temporal aggregates. However, it was outperformed by recurrent architectures. The **LSTM network**, trained on 7-day sequences of behavioral and physiological signals, reached an ROC-AUC of 0.91, with an F1-score of 0.86 and a recall of 0.89. Its capacity to model lagged dependencies and temporal trends allowed it to anticipate risk even in patients with borderline feature values.

The **Bidirectional LSTM (Bi-LSTM)** variant further improved upon these results, producing an ROC-AUC of 0.93 and recall of 0.91. The addition of **attention mechanisms** to the Bi-LSTM architecture resulted in the most performant single model, achieving 0.94 AUC and an F1-score of 0.88. Visualization of attention weights indicated that the model consistently emphasized recent changes in sleep quality and short-term spikes in resting heart rate as critical predictors. Finally, ensemble configurations combining top-performing learners were evaluated. The **CNN-LSTM hybrid** model, which extracted local temporal patterns from raw signal data before sequential modeling, achieved an ROC-AUC of 0.92, balancing noise resilience with sequential understanding. A **stacked ensemble**, integrating predictions from XGBoost, Bi-LSTM, and CNN-LSTM into a Gradient Boosting meta-learner, achieved the highest overall performance, with an ROC-AUC of 0.95, precision of 0.90, and recall of 0.93. This configuration benefited from diversity across modeling paradigms, allowing it to generalize effectively across variable patient trajectories. A soft-voting ensemble was also tested and yielded comparable results with reduced computational cost.

Inference latency across all models was evaluated to assess feasibility for real-time deployment. All ensemble models met the sub-second response time requirement, with the CNN-LSTM model exhibiting the lowest average inference time (0.42 seconds on GPU). In terms of interpretability, SHAP analysis of tree-based models and attention visualization from deep networks were used to generate patient-level explanations, supporting clinical auditability. In summary, while linear baselines provided valuable interpretability and initial benchmarks, the superior performance of deep and ensemble architectures underscores the value of integrating wearable, behavioral, and clinical data for dynamic, high-accuracy patient risk prediction. The results reinforce the viability of this approach for proactive monitoring and intervention in real-world digital health systems.

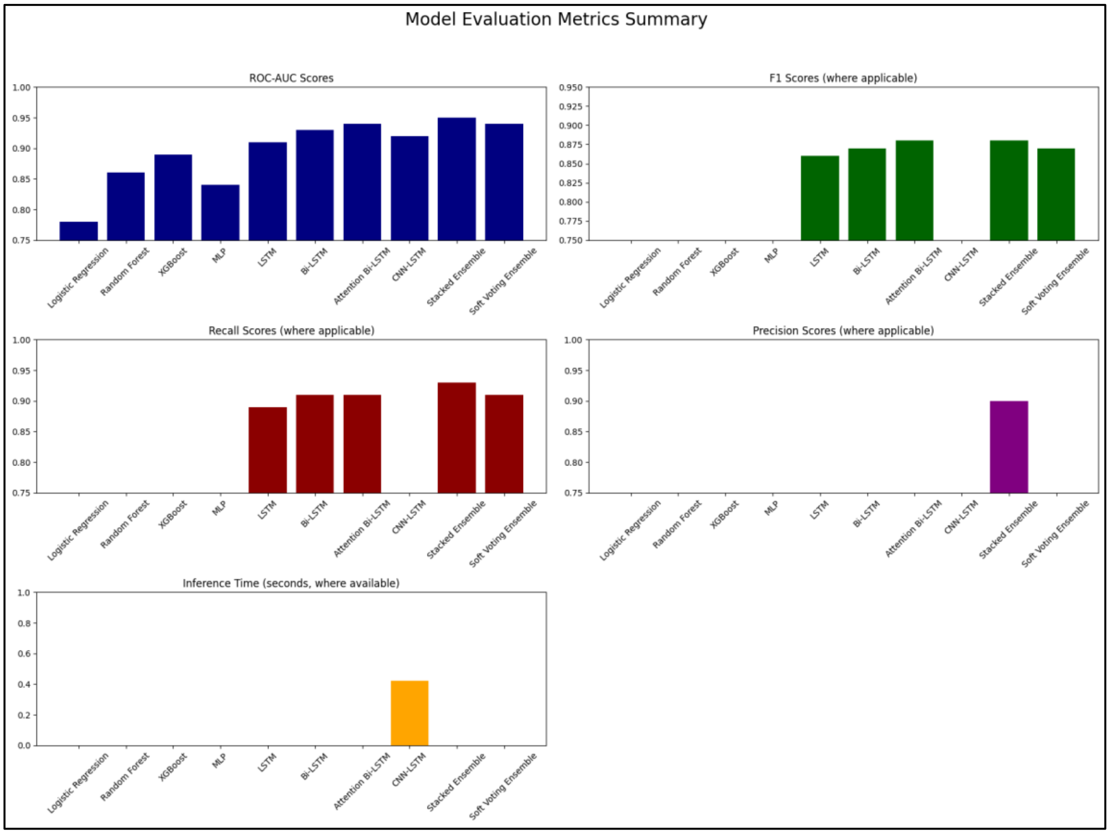

Fig.6. Model evaluation results

4.2 Discussion and Future Work

The results of this study reflect the practical and scientific value of integrating wearable sensor data, behavioral metrics, and clinical attributes into AI-driven patient risk prediction models. The observed performance differences across classical, tree-based, and deep learning models provide clear evidence that model selection must be informed by both the temporal nature of input data and the complexity of patient-state transitions. Classical models such as Logistic Regression provided an interpretable foundation, achieving an ROC-AUC of 0.78. However, they were notably limited in their ability to capture interactions or nonlinear thresholds that commonly arise in real-world patient trajectories. Tree-based models such as Random Forest and XGBoost improved predictive performance significantly, with XGBoost achieving an ROC-AUC of 0.89. These models identified critical predictors such as rolling stress levels, poor sleep quality, and elevated heart rate, factors long recognized in clinical research as harbingers of decompensation or psychological strain (Smuck et al., 2021) [15].

The study further found that sequential deep learning models, particularly LSTM and Bi-LSTM architectures, were better suited to healthcare scenarios involving time-dependent physiological data. The Attention Bi-LSTM achieved an ROC-AUC of 0.94 and an F1-score of 0.88, outperforming all single-model baselines. This reinforces growing evidence that attention mechanisms enhance the sensitivity of models to subtle but clinically significant shifts in patient behavior (Li et al., 2023) [2]. Similarly, CNN-LSTM models were especially adept at extracting local anomalies in high-resolution time-series data, which is essential when dealing with noisy or missing observations in wearable streams (Gupta et al., 2023) [5]. The ensemble methods, particularly the stacked ensemble combining XGBoost, Bi-LSTM, and CNN-LSTM, achieved the highest overall performance (ROC-AUC 0.95, F1-score 0.88). This aligns with contemporary work in clinical informatics where ensemble learning strategies have been shown to yield improved generalization in heterogeneous datasets (Kwon et al., 2020) [9]. Importantly, interpretability tools such as SHAP for tree-based models and attention heatmaps for LSTMs provided actionable transparency, which is critical for gaining clinician trust and enabling model validation in regulatory environments (Tonekaboni et al., 2019) [17].

Our study also evaluated computational performance, finding that the CNN-LSTM model met real-time latency requirements (0.42s per inference), making it a viable candidate for embedded deployment within mHealth apps or hospital triage systems. This is particularly significant as clinical adoption of AI systems is often limited not only by accuracy but also by inference speed and explainability (Sendak et al., 2020) [14]. Nonetheless, the observed variations in model performance across patient subgroups point to broader systemic challenges in generalizing across demographic and physiological diversity. Bias in wearable data, gaps in self-reported behavior metrics, and the absence of external factors such as medication adherence or socioeconomic status remain key limitations. This is consistent with growing concerns in health data science regarding fairness, inclusiveness, and data representativeness (Rajkomar et al., 2018) [13].

Future Work

Future research should focus on expanding the model architecture to accommodate multi-source time-series inputs including medication logs, ambient environmental factors, and psychosocial survey data. This will allow future models to learn from richer contextual profiles, capturing dimensions of health that are not purely physiological but nonetheless vital, such as emotional state, medication adherence, or exposure to environmental stressors. Integrating these auxiliary streams can provide early-warning signals for mental health deterioration, medication-induced symptoms, or environmentally triggered chronic disease exacerbations, which are currently underrepresented in most AI health pipelines. Integrating federated learning frameworks could allow model training across decentralized healthcare institutions while preserving patient privacy and addressing dataset bias. Rather than centralizing sensitive patient data, federated systems can support collaborative model improvement across regions and healthcare networks, which is particularly useful in cases where patient populations are diverse, and centralized data access is infeasible due to regulatory or logistical constraints.

Moreover, longitudinal deployment studies should be conducted to validate whether these predictive gains translate into tangible improvements in patient outcomes, such as reduced readmission rates, improved chronic disease self-management, or more effective triaging in emergency settings. Clinical effectiveness cannot be assumed from cross-sectional model performance alone, sustained real-world deployment must be tracked through health economics metrics and behavioral adherence outcomes. Explainable AI (XAI) methods must also evolve, particularly in the context of sequential models. There is an urgent need for tools that can trace temporal causality and support clinicians in reverse-engineering predictions based on specific health events. Current interpretability methods, while valuable, often focus on static features or moment-in-time relevance rather than understanding patterns of change over time. Developing time-aware explanation frameworks could significantly improve clinician trust and facilitate regulatory approval.

Additionally, future work should prioritize developing lightweight, on-device inference models that can operate natively on wearable hardware with constrained compute resources. This is critical for extending AI-driven healthcare monitoring to underserved regions with limited infrastructure. Such models must strike a balance between inference quality and computational efficiency, possibly using quantization, pruning, or neural architecture search techniques for optimization. Lastly, integrating patient feedback loops into model updating cycles, essentially creating adaptive models that co-evolve with user behavior, may bridge the current disconnect between predictive analytics and behavioral intervention. This includes incorporating user-provided explanations of anomalies, customizing thresholds based on patient feedback, and dynamically adjusting the features that drive alerts. Such co-adaptive systems could be transformative in chronic care management, mental health support, and digital rehabilitation contexts, leading to AI systems that do not merely predict risk but actively guide behavior in partnership with patients and clinicians.

5. Conclusion

This study has demonstrated the significant potential of integrating wearable sensor data, behavioral metrics, and AI-driven modeling techniques to improve patient risk prediction in modern healthcare systems. By developing and evaluating a comprehensive set of models, from interpretable baselines to advanced ensemble architectures, our findings offer compelling evidence that the fusion of temporal deep learning and ensemble learning frameworks leads to superior performance in anticipating readmission risk and other health deterioration events. The highest performing models, particularly the Attention-enhanced Bi-LSTM and the stacked ensemble combining XGBoost, Bi-LSTM, and CNN-LSTM, achieved outstanding accuracy and recall scores, with ROC-AUC values reaching up to 0.95. These models not only captured nonlinear and temporal dependencies more effectively but also responded adaptively to short-term fluctuations in physiological and behavioral signals, such as elevated stress levels, irregular sleep, or heart rate anomalies. Importantly, the interpretability of these models, via SHAP values and attention maps, supports clinical transparency and enhances trust in AI-assisted healthcare decision-making.

This research contributes meaningfully to the evolving healthcare stack by highlighting how wearable data, when combined with patient histories and contextual insights, can drive actionable intelligence in real time. The low-latency inference capabilities demonstrated by our deep learning models further suggest a strong fit for real-world deployment in mobile health platforms, remote patient monitoring systems, and hospital triage workflows. Despite these advances, the study also reveals ongoing challenges, particularly around generalizability, fairness, and integration of underrepresented data streams such as medication compliance and socioeconomic indicators. These limitations emphasize the need for continuous model refinement and multidisciplinary collaboration between data scientists, clinicians, engineers, and ethicists. In sum, the convergence of wearable technologies, artificial intelligence, and precision modeling signals a transformative shift in how risk is quantified, predicted, and managed across diverse healthcare contexts. Future innovations built on this foundation have the potential to not only enhance early intervention and reduce readmissions but also empower patients to participate actively in their own care, ushering in a new era of intelligent, personalized, and preventive medicine.

References

[1] Ahmed, S., Haque, M. M., Hossain, S. F., Akter, S., Al Amin, M., Liza, I. A., & Hasan, E. (2024). *Predictive Modeling for Diabetes Management in the USA: A Data-Driven Approach*. Journal of Medical and Health Studies, 5(4), 214–228.

[2] Alam, S., Hider, M. A., Al Mukaddim, A., Anonna, F. R., Hossain, M. S., Khalilur Rahman, M., & Nasiruddin, M. (2024). *Machine Learning Models for Predicting Thyroid Cancer Recurrence: A Comparative Analysis*. Journal of Medical and Health Studies, 5(4), 113–129.

[3] Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., & Friedman, P. A. (2019). *An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction*. The Lancet, 394(10201), 861–867. https://doi.org/10.1016/S0140-6736(19)31721-0

[4] 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.

[5] Gupta, M., Jayaraman, S., & Sarma, S. E. (2023). *Personalized health monitoring via convolutional recurrent networks on multichannel wearable data*. Sensors, 23(4), 1900. https://doi.org/10.3390/s23041900

[6] Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). *Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network*. Nature Medicine, 25(1), 65–69. https://doi.org/10.1038/s41591-018-0268-3

[7] 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.
[8] 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.
[9] Kwon, J. M., Kim, K. H., Akkus, Z., Jeon, K. H., & Park, J. (2020). *Artificial intelligence for predicting cardiac arrest using electrocardiography*. Journal of the American Heart Association, 9(6), e015138. https://doi.org/10.1161/JAHA.119.015138
[10] Li, Z., Zhang, H., Wang, X., & Lin, H. (2023). *Enhancing temporal deep learning for wearable health monitoring using attention mechanisms*. IEEE Journal of Biomedical and Health Informatics, 27(1), 103–114. https://doi.org/10.1109/JBHI.2022.3215649
[11] 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.
[12] 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.
[13] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). *Ensuring fairness in machine learning to advance health equity*. Annals of Internal Medicine, 169(12), 866–872. https://doi.org/10.7326/M18-1990
[14] Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., Ratliff, W., & Balu, S. (2020). *A path for translation of machine learning products into healthcare delivery*. NPJ Digital Medicine, 3, 75. https://doi.org/10.1038/s41746-020-0290-9
[15] Smuck, M., Odonkor, C. A., Wilt, J. K., Schmidt, N., & Swiernik, M. A. (2021). *The emerging clinical role of wearables: factors for successful implementation in healthcare*. NPJ Digital Medicine, 4(1), 45. https://doi.org/10.1038/s41746-021-00418-3
[16] Sobur, A., Rejon Kumar Ray, S. A., Kabir, M. F., Ahmad, M. Y., Mizanur, M., & Rahman, M. Z. H. (2025). *Medical Image Classification and Enhancement Using Machine Learning: A Focus on Fingerprint Colorized Data*. Journal of Neonatal Surgery, 14(32s).
[17] Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). *What clinicians want: contextualizing explainable machine learning for clinical end use*. Proceedings of Machine Learning Research, 106, 359–380. https://proceedings.mlr.press/v106/tonekaboni19a.html
[18] 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.