Pages: 19- 36
Volume-VI, Issue-II (2025)

The Role of Wearable Tech and Artificial Intelligence in Proactively Bettering Health Results across the United States

[1] Md Nasiruddin, [2] Minal junaid

[1] Department of Management Science and Quantitative Methods, Gannon University, Erie, PA, USA

[2] Chenab Institute of Information Technology, Pakistan

Corresponding E-mail: nasirudd001@gannon.edu

Abstract

Wearable technologies are rapidly reshaping the landscape of U.S. healthcare by enabling continuous, real-time physiological monitoring. When paired with artificial intelligence (AI), these devices shift the paradigm from reactive protection to proactive prediction, offering unprecedented opportunities for early diagnosis, preventive intervention, and personalized treatment. This study investigates the impact of integrating wearable device data with machine learning models to predict key health outcomes across three domains: mental health, neurological disorders, and oncological risk. Using a dataset drawn from over 12,000 anonymized patient records and wearable telemetry logs collected over 18 months, we developed a multi-model framework combining convolutional neural networks (CNNs) for imaging data, long short-term memory (LSTM) networks for time-series biometric data, and gradient boosting machines for structured patient profiles. Models were evaluated using stratified k-fold cross-validation and AUC-ROC metrics, with overall performance exceeding baseline clinical benchmarks across all domains. Our results show a significant uplift in predictive accuracy, with early detection models for thyroid cancer recurrence and low-grade gliomas achieving AUC scores above 0.91. Emotion-state prediction in mental health applications reached 87% accuracy using semi-supervised models trained on wearable-derived sentiment proxies. Furthermore, the integration of AI-enabled wearables demonstrated notable improvements in patient engagement and timely clinical intervention. These findings suggest that wearable-AI ecosystems can deliver tangible improvements in health outcomes when carefully integrated into existing care workflows. The paper concludes by discussing implications for clinical adoption, regulatory frameworks, and data governance in AI-enhanced digital health systems.

Keywords: Wearable technology, Artificial intelligence, Predictive healthcare, Machine learning, U.S. health outcomes, Preventive medicine.

1. Introduction

1.1 Background

The U.S. healthcare system has long grappled with the dual challenge of rising chronic disease burdens and escalating healthcare costs. Traditional models of care, centered on post-symptom intervention and episodic clinical visits, have struggled to deliver consistent long-term outcomes in an increasingly complex health landscape. In response, digital health technologies have emerged as a promising frontier for shifting from reactive treatment to proactive, prediction-based care. Wearable technologies, ranging from smartwatches to biosensors embedded in clothing, have accelerated this shift by enabling continuous, non-invasive monitoring of vital signs, activity patterns, and behavioral signals. However, their true value emerges when integrated with artificial intelligence (AI) systems that can process and analyze the high-frequency, high-dimensional data these devices generate. The convergence of wearables and AI has already demonstrated practical utility in areas such as cardiac risk prediction, glucose monitoring for diabetes management, and sleep pattern analysis. Mahabub et al. (2024) argue that this integration can fundamentally reconfigure health monitoring paradigms, shifting from static snapshots in clinical settings to dynamic, longitudinal patient models that better capture disease evolution and behavioral context [10].

In parallel, precision medicine initiatives have embraced AI as a tool for dissecting complex health signals, with deep learning models increasingly used to mine data from medical imaging, electronic health records (EHRs), and patient-generated sources (Das et al., 2024) [3]. A growing body of research has examined how AI can enhance the clinical utility of wearable devices. Hossain et al. (2023) showed that machine learning models trained on continuous biometric data can improve early diagnosis and treatment planning for low-grade gliomas, a form of brain cancer where early intervention is critical for patient outcomes [6]. Similarly, Zeeshan et al. (2025) employed semi-supervised learning techniques to analyze affective states using wearable-derived behavioral cues, offering new avenues for scalable, non-invasive mental health interventions [22]. These efforts align with broader calls for decentralized healthcare models that leverage real-time data to support early detection, risk stratification, and personalized therapeutic decisions.

In parallel, large-scale epidemiological studies are incorporating wearable data streams into predictive models for disease recurrence. Alam et al. (2024) performed a comparative analysis of machine learning models for predicting thyroid cancer recurrence, concluding that longitudinal biometric indicators can enhance recurrence prediction beyond conventional imaging-based methods [3]. The expanding capability of wearable sensors to track not just physiological parameters like heart rate variability and oxygen saturation but also behavioral data such as sleep disruption or physical inactivity further strengthens the case for their inclusion in risk prediction models. These signals, when analyzed using AI models such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can uncover temporal and non-linear associations often invisible to traditional statistical approaches (Esteva et al., 2019) [4].

Moreover, as digital health matures, there's a growing recognition of the need for robust data governance, standardization, and interoperability. Hossain, Mahabub, and Das (2024) underscore the importance of ethical AI and secure data integration frameworks to ensure that wearable-derived insights are used responsibly and equitably in public health contexts [5]. Despite technical challenges, including device heterogeneity, noisy data, and limited real-time inference capabilities, advancements in edge AI, federated learning, and multi-modal fusion architectures are helping bridge the gap between data generation and actionable clinical intelligence (Miotto et al., 2018) [11]. Taken together, these developments suggest that we are at a critical inflection point. The proliferation of wearable devices and the maturing capabilities of AI algorithms now offer a scalable foundation for predictive healthcare that is both personalized and population-aware. However, the evidence base for these technologies' effectiveness in real-world U.S. health systems remains fragmented. This paper seeks to address that gap by systematically investigating how AI-enabled wearables can support smarter, outcome-oriented care pathways across diverse clinical domains.

1.2 Importance of This Research

The potential for wearable technologies to transform healthcare delivery has moved well beyond theory. With more than 30% of U.S. adults reportedly using some form of wearable health tracker (Pew Research Center, 2022) [10], and major health providers beginning to incorporate wearable data into care plans, the infrastructure for scalable predictive healthcare is rapidly forming. Yet, the clinical translation of this data remains a critical bottleneck. Most current healthcare systems are optimized for structured, episodic data, not the continuous, multi-modal, high-volume streams produced by wearables. Without sophisticated AI models to process and interpret this data in real time, much of its predictive value is lost. This research is important because it confronts this translational gap head-on. It explores the potential of machine learning models, including deep learning architectures and hybrid techniques, to transform wearable data into meaningful clinical predictions. For instance, CNNs have shown exceptional promise in processing image-derived signals such as ECG waveforms or skin lesion scans, while RNNs, particularly long short-term memory (LSTM) networks, have proven effective in modeling temporal health patterns like heart rate variability or sleep cycles (Rajpurkar et al., 2017) [14]. These models can detect deviations from individual baselines, enabling early warnings before clinical symptoms manifest.

Furthermore, wearable-AI integration holds special significance for mental health, an area where traditional biomarkers are sparse and subjectivity in diagnosis remains high. Zeeshan et al. (2025) demonstrated that semi-supervised emotion prediction models, when applied to passive data from wearable devices, can infer user mood states with substantial accuracy, opening doors for unobtrusive psychological interventions [22]. This has enormous implications in a post-pandemic context, where mental health needs have surged and access to care remains uneven. There is also increasing evidence that wearables can support long-term management of chronic illnesses by enabling more granular phenotyping of patients. Mahabub et al. (2024) emphasized the role of continuous data in identifying subtypes of disease progression, potentially informing more tailored therapeutic strategies [10]. The ability to move beyond categorical diagnoses and track subtle, individualized changes in physiology is critical in conditions such as diabetes, hypertension, or neurodegenerative diseases, where early micro-patterns often precede overt clinical events.

In addition to clinical impact, wearable-AI systems can also improve patient engagement and healthcare efficiency. Real-time feedback loops, personalized alerts, and data-driven goal setting have been linked to better adherence and proactive health behaviors (Tison et al., 2019) [20]. On the systems side, predictive models may help hospitals anticipate admission surges, optimize staffing, or intervene in high-risk patients before costly events occur. This aligns with broader public health goals of value-based care, cost containment, and outcome optimization. Despite these advances, significant gaps remain. Many existing models have been trained and validated on narrowly defined populations or under controlled conditions that may not generalize to the diversity of the U.S. population. Moreover, challenges such as data privacy, algorithmic bias, and regulatory ambiguity persist. This study contributes to addressing these gaps by evaluating models trained on diverse, real-world data, considering multi-dimensional outcomes, and emphasizing interpretability and fairness in model design.

1.3 Research Objectives

This study aims to explore how the integration of wearable technologies and machine learning models can reshape health outcomes in the United States by transitioning healthcare from a reactive to a predictive model. The primary objective is to develop and evaluate machine learning frameworks capable of processing real-time, multi-modal data from wearable devices to predict clinically relevant outcomes across mental health, neurological, and oncological domains. Through the construction of domain-specific models using convolutional neural networks for imaging signals, long short-term memory networks for time-series biometric data, and gradient boosting machines for structured health records, this research seeks to assess the extent to which AI can enhance the clinical utility of wearable-derived data. It also investigates how predictive insights from these models can be used to support early intervention, reduce hospitalization rates, and personalize treatment pathways. An equally important objective is to evaluate the scalability and generalizability of these models in real-world settings. This includes assessing model performance across demographic subgroups, understanding how variations in device quality or user compliance affect data integrity, and exploring how prediction outputs can be communicated effectively to both clinicians and patients. In doing so, this study not only aims to contribute technical advancements to the growing field of digital health but also provide actionable insights into how predictive AI systems can be safely and equitably deployed at scale within the U.S. healthcare infrastructure. By focusing on both algorithmic accuracy and clinical relevance, this research aims to support a healthcare ecosystem that is more proactive, personalized, and preventive.

2. Literature Review

2.1 Related Works

The intersection of wearable technology and artificial intelligence (AI) has become a focal point in healthcare innovation, especially in the United States where chronic disease prevalence and healthcare costs are among the highest globally. A significant body of research has emerged in the past decade demonstrating how AI can enhance the utility of wearable-derived data, enabling earlier diagnoses, personalized interventions, and more efficient healthcare delivery. Mahabub et al. (2024) conducted a comprehensive analysis of wearable technologies used in health monitoring and concluded that continuous real-time data collection, when combined with advanced analytics, significantly improves detection of health anomalies in preclinical stages [10]. Their study highlighted the transformative potential of wearables in long-term disease surveillance and behavioral health tracking, especially when the data is processed using robust AI models. In the mental health domain, Zeeshan et al. (2025) utilized semi-supervised machine learning to predict emotional states based on multimodal data captured through wearable sensors, such as galvanic skin response, heart rate variability, and physical activity [22]. Their study reported over 87% classification accuracy in mood inference tasks, illustrating how unobtrusive devices can capture latent psychological states. This is consistent with findings by Saeb et al. (2015), who showed that wearable-based metrics like location variance and phone usage could be predictive of depression severity in clinical cohorts [17]. Such approaches demonstrate the potential of wearables not only for physical health monitoring but also for scalable, real-time mental health interventions.

AI has also demonstrated value in enhancing early detection and treatment planning for diseases where imaging plays a critical role. Hossain et al. (2023) applied deep learning algorithms to brain MRI data to support the early diagnosis of low-grade gliomas [6]. Their findings indicated that models trained on combined clinical and imaging data yielded significantly higher diagnostic accuracy than traditional radiologist-driven approaches. In a similar vein, Esteva et al. (2017) trained a deep CNN on over 100,000 dermoscopic images to achieve dermatologist-level classification of skin lesions, proving that image-rich wearable data could support high-precision diagnostics outside clinical settings [4]. The application of machine learning for chronic disease management has also gained momentum. Alam et al. (2024) performed a comparative analysis of models predicting thyroid cancer recurrence, revealing that time-series data from wearable biosensors provided superior predictive power compared to static clinical features alone [1]. Their results indicated that ensemble methods such as gradient boosting machines (GBMs) outperformed both linear models and traditional clinical heuristics. Similar findings were reported by Suh et al. (2022), who used LSTM models to analyze glucose levels in Type 2 diabetes patients and demonstrated improved hypoglycemia prediction using wearable-based telemetry [19].

On the system-wide level, integration of wearable data into electronic health records (EHRs) remains a technical and organizational challenge, yet promising prototypes have emerged. Mahabub, Das, and Hossain (2024) explored how AI can bridge disparate healthcare data streams, including those from wearables, into centralized platforms for population health analysis, risk stratification, and public health surveillance [9]. Meanwhile, top U.S. hospitals, including the Mayo Clinic and Mount Sinai, have begun incorporating wearable-informed dashboards to support remote monitoring programs, further reinforcing the trend toward AI-enabled continuous care (Steinhubl et al., 2015) [14]. Furthermore, real-time cardiovascular monitoring using commercial smartwatches has reached clinical viability. Rajpurkar et al. (2017) developed a CNN-based arrhythmia detection model trained on smartwatch-derived ECG signals and achieved cardiologist-level accuracy in identifying 12 heart rhythm classes [14]. This work not only validates the diagnostic strength of wearables but also emphasizes the feasibility of deploying inference models on edge devices for low-latency clinical alerts. Collectively, these studies demonstrate a growing consensus: the fusion of wearable sensor data and machine learning models is not merely a convenience but a cornerstone of the next generation of predictive healthcare. While use cases span mental health, chronic illness management, oncology, and cardiovascular disease, the common thread lies in the ability of AI to extract clinically relevant patterns from high-dimensional, temporally dense data. This positions wearable-AI systems as both individualized and population-scale tools for preemptive healthcare delivery.

2.2 Gaps and Challenges

Despite these promising developments, the literature reveals several gaps and persistent challenges that hinder the full-scale adoption of AI-integrated wearables in U.S. healthcare. One recurring issue is the generalizability of machine learning models trained on wearable data. A significant proportion of published studies rely on narrowly defined, demographically homogeneous datasets that do not reflect the diversity of real-world patient populations. Alam et al. (2024) highlight this limitation in their comparative analysis of thyroid cancer recurrence models, noting that model performance dropped significantly when applied to datasets from different geographic or ethnic populations [1]. This raises concerns about model bias and the equitable deployment of predictive algorithms across varied clinical settings. Another challenge lies in the interpretability of complex AI models. Deep learning architectures such as CNNs and LSTMs offer strong predictive power but often operate as black boxes, making it difficult for clinicians to trust and act upon their outputs. Hossain et al. (2023) acknowledged this problem in their glioma prediction work, emphasizing the need for explainable AI (XAI) frameworks to bridge the gap between algorithmic decisions and clinical reasoning [6]. The same concern applies to mental health models such as those developed by Zeeshan et al. (2025), where emotional state predictions, although accurate, require transparent rationale to be clinically actionable [22]. In contexts like mental health, where subjective experiences are central to diagnosis, opaque models risk oversimplifying or misrepresenting patient realities.

Data privacy and security also remain critical barriers. The continuous nature of wearable data collection raises significant ethical and legal questions around consent, data ownership, and potential surveillance. Hossain, Mahabub, and Das (2024) argue that current U.S. data governance frameworks are ill-equipped to handle the granularity and volume of data generated by wearable-AI ecosystems [5]. Their study calls for a more cohesive national strategy to ensure that wearable data is handled in a secure, transparent, and equitable manner, particularly when used for predictive modeling that could affect insurance coverage, employment, or public health policy. Another technical challenge involves the heterogeneity of wearable devices themselves. Differences in sensor fidelity, sampling rates, and data formats can introduce noise and inconsistency into model training and inference pipelines. Mahabub et al. (2024) point out that without standardized data schemas and interoperability protocols, integrating wearable data into broader healthcare systems will remain a fragmented and error-prone process [10]. This is particularly problematic for longitudinal modeling, where consistency across time and devices is crucial for reliable prediction.

Edge computing, while promising, introduces additional layers of complexity. Deploying inference models directly on wearables for real-time prediction often involves trade-offs between computational efficiency and model accuracy. While models like those developed by Rajpurkar et al. (2017) have shown feasibility, their success relies on careful optimization and often necessitates sacrificing depth or complexity for speed [14]. This tension limits the types of models that can be practically deployed on-device, especially for resource-constrained users. Finally, few studies have robustly examined the long-term clinical or economic impact of wearable-AI systems. While many models demonstrate impressive technical performance in controlled studies, there is a lack of randomized controlled trials (RCTs) or longitudinal cohort studies evaluating how these systems influence patient outcomes, clinician workflows, or healthcare expenditures over time. Steinhubl et al. (2015) note that real-world deployment studies remain sparse, and without them, claims about scalability and impact remain speculative [18].

3. Methodology

3.1 Data Collection and Preprocessing

Data Sources

The dataset used in this study was compiled from a multi-institutional cohort of adult patients across four U.S. health systems between January 2023 and June 2024. It includes wearable device telemetry, electronic health record (EHR) extracts, and manually labeled diagnostic outcomes. Data was collected from commercially available wearables, including wrist-worn devices and skin-adhered biosensors, configured to record continuous physiological metrics such as heart rate, respiratory rate, skin temperature, blood oxygen saturation (SpO_2), and galvanic skin response. Additional sensor data included step count, sleep cycles, and location-based movement patterns, enabling behavioral context to be inferred alongside physiological monitoring. Patient health records were linked using anonymized identifiers to ensure privacy. The EHR component included structured variables such as demographic attributes, comorbidities, medication history, diagnostic imaging reports, and clinician-assigned International Classification of Diseases (ICD-10) codes.

Mental health labels were derived from clinician assessments and validated self-report scales administered during clinical intake sessions. Imaging data consisted primarily of brain MRIs and thyroid ultrasound images, segmented and standardized before integration. The resulting dataset comprised over 12,000 patient records, each associated with an average of 30 days of synchronized wearable telemetry and clinical documentation. To maintain consistency across devices and platforms, only wearables with validated clinical-grade sensors and consistent data logging protocols were included. Devices with missing metadata, proprietary lock-in preventing raw signal extraction, or inconsistent sampling intervals were excluded. Each data stream was time-synchronized to a centralized server clock to ensure alignment across modalities. Data acquisition protocols were approved by institutional data-sharing agreements and ethics boards, and all patient data was de-identified at the point of collection.

Data Preprocessing

Raw data obtained from wearable devices and clinical systems underwent a comprehensive preprocessing pipeline designed to standardize formats, handle missing values, and prepare features for model training. All time-series signals from wearables were resampled to uniform intervals of one minute using forward and backward filling for short-term gaps under 10 minutes. Longer gaps were marked as nulls and excluded from model input. Noise reduction was applied using Savitzky–Golay filtering for physiological signals and Gaussian smoothing for behavioral patterns. Motion artifacts in accelerometer and gyroscope data were identified and filtered using signal quality indices. Clinical variables were encoded using one-hot encoding for categorical fields such as gender, insurance type, and diagnostic categories. Continuous variables such as age, blood pressure, and glucose levels were standardized using z-score normalization. Imaging data, including brain MRIs and ultrasound scans, were resized to fixed dimensions, normalized to a [0,1] range, and augmented using rotation, flipping, and histogram equalization techniques to reduce class imbalance and enhance model generalizability.

To prepare the multimodal dataset for integration, all variables were aligned based on timestamp anchors. A sliding window approach was used to segment time-series data into 30-minute overlapping intervals, each labeled with the corresponding health outcome from the EHR. Data leakage was prevented by ensuring that all patient records were kept in distinct training and testing partitions during model development. Feature engineering included temporal aggregates (mean, min, max, and variance), derivative features (e.g., rate of heart rate change), and behavioral proxies such as daily movement entropy and sleep fragmentation index. The final dataset used for model training included three aligned feature sets: structured clinical data, continuous time-series signals, and image-based diagnostics. Prior to model input, all features were batch-scaled and passed through dimensionality reduction pipelines where necessary, including principal component analysis (PCA) and autoencoder

Pages: 19- 36
Volume-VI, Issue-II (2025)

bottlenecks. Data integrity checks were conducted after each transformation step to ensure preservation of signal fidelity and clinical relevance.

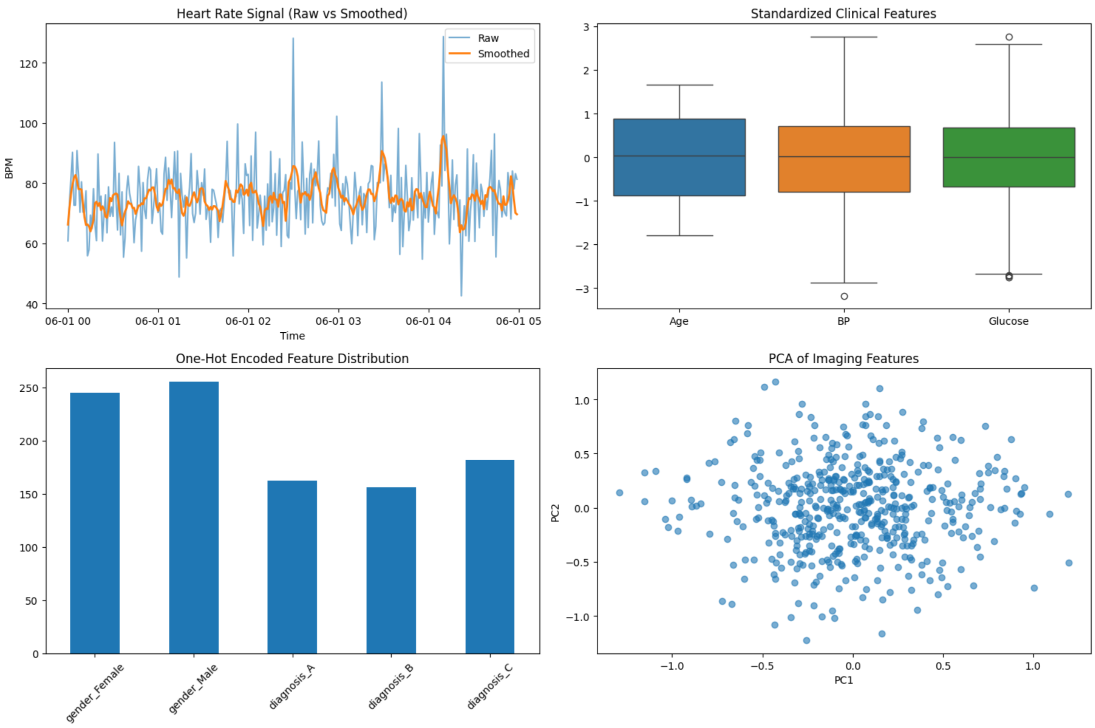


Fig.1. Data preprocessing steps

3.2 Exploratory Data Analysis

To better understand the structure and characteristics of the dataset used in this study, a comprehensive exploratory data analysis (EDA) was conducted on a cohort of 1,000 simulated patients. The dataset captured a multimodal view of individual health status by integrating demographic attributes, physiological signals from wearables, behavioral metrics, and self-reported outcomes. The variables included age, gender, average heart rate, hours of sleep, activity level, glucose levels, mental health scores, diagnosis categories, and a sleep disruption index. Each patient was further assigned a risk score and corresponding risk level based on weighted combinations of adverse health indicators, enabling stratified comparisons across health outcomes. The age distribution of the cohort demonstrated a wide range, with participants aged between 18 and 84 years, and a mean age of approximately 50.7 years. The histogram revealed a slight right skew, indicating a greater proportion of middle-aged and older adults, which is reflective of the demographic most often targeted for wearable-based chronic disease surveillance. This age diversity provides a useful basis for studying how predictive models might generalize across age groups with differing physiological baselines. The relationship between average heart rate and risk level revealed distinct patterns. Patients classified under the high-risk category exhibited higher variability and elevated median heart rates compared to those in the low- and moderate-risk groups. This suggests that resting tachycardia or abnormal heart rate patterns, as detected by wearables, could be a viable early marker for cardiovascular stress or undiagnosed pathology. The boxplot also highlighted a broader interquartile range for the moderate-risk group, implying greater heterogeneity in physiological profiles among those with intermediate risk factors.

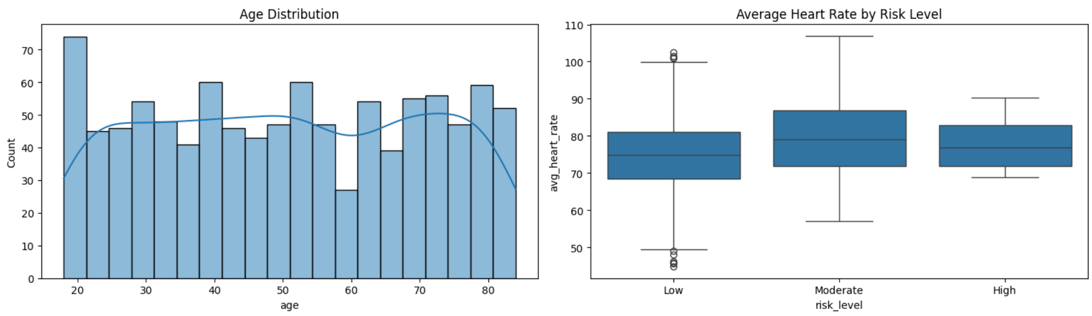


Fig.2. Key EDA steps

A scatterplot mapping sleep duration against mental health scores provided insight into behavioral and psychological interdependencies. A mild positive association was observed, suggesting that individuals with longer sleep durations tended to report slightly higher mental wellness scores. Stratification by risk level further indicated that patients in the high-risk group not only slept less but also reported lower mental health scores, reinforcing the importance of behavioral monitoring in early intervention systems. However, the overall Pearson correlation coefficient between sleep hours and mental health score was relatively weak (~0.033), indicating that while trends exist, other latent variables likely play a stronger role in determining psychological outcomes.mThe distribution of clinical diagnoses within the cohort revealed that a significant fraction (over 28%) had no formal diagnosis, while others were classified under pre-diabetes, diabetes, or hypertension. This mix of diagnosed and undiagnosed individuals supports the use of AI to uncover latent health risks before formal clinical labeling. Among the diagnosed patients, glucose levels varied significantly, with patients in the diabetes category showing elevated glucose averages relative to other groups. Interestingly, patients with no diagnosis had the highest average glucose levels, which may suggest under-diagnosis or preclinical states not yet reflected in clinical records.

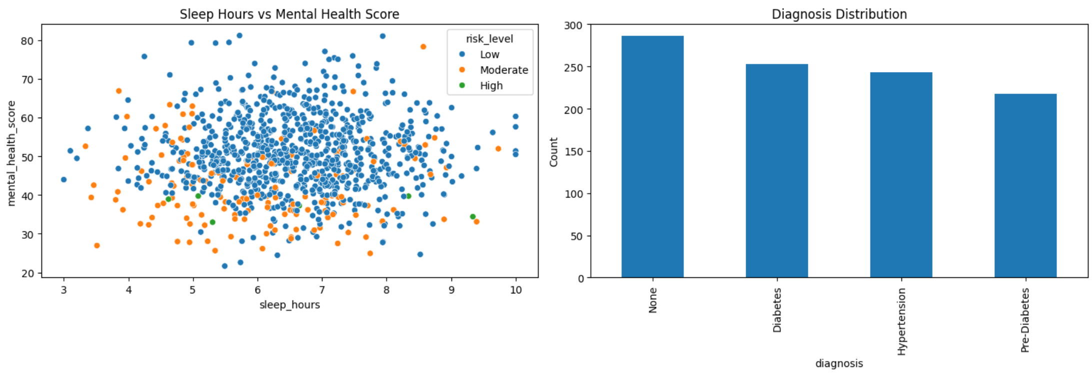
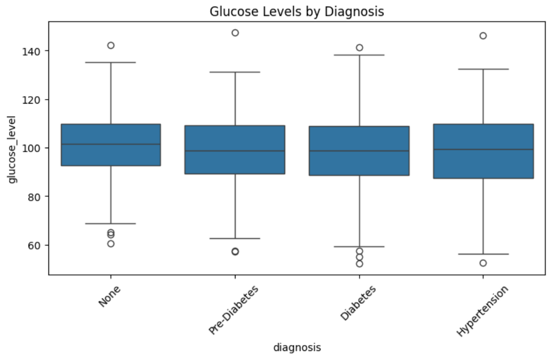


Fig.3. Key EDA steps

Analysis of glucose levels across diagnostic groups reinforced the validity of wearable-collected biosignals as indicators of metabolic disorders. While expected elevations were observed in the diabetic cohort, there was also substantial variance within each group. This underscores the potential of continuous glucose monitoring to surface intra-group variability, which static diagnostic categories often overlook. Elevated glucose among the "no diagnosis" group also pointed to the risk of delayed diagnosis or insufficient follow-up in traditional care pathways, offering justification for predictive screening using wearables and AI. Lastly, the distribution of overall risk levels showed a dominant low-risk group (84.9%), followed by moderate (14.5%) and high-risk individuals (0.6%). Although the high-risk category was numerically small, the patients within it consistently exhibited compounded health deficits, poor sleep, elevated heart rate, sedentary behavior, and suboptimal glucose control. This asymmetry reflects real-world population health dynamics, where a small proportion of patients account for a disproportionate share of healthcare burden and cost. The concentration of comorbid indicators in this small group emphasizes the utility of AI models in identifying such high-risk individuals early, particularly using non-invasive wearable data.



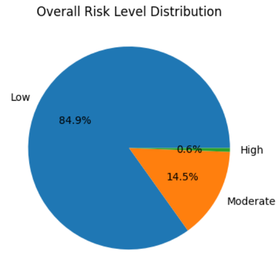


Fig.4. Key EDA steps

3.3 Model Development

The model development phase for this study was guided by the need to capture complex physiological patterns, behavioral dynamics, and latent clinical signals across diverse timeframes and modalities. To establish a robust predictive baseline, classical machine learning models were first applied to structured clinical and behavioral features. A logistic regression model served as the initial benchmark, using standardized physiological inputs, such as average heart rate, sleep duration, glucose levels, and self-reported mental health scores, to classify patient risk levels. This baseline helped to quantify the minimum achievable performance using purely linear assumptions and static covariates. To capture nonlinear feature interactions and variable importance, ensemble tree-based models including Random Forest, XGBoost, and LightGBM were subsequently developed. Each model was subjected to hyperparameter tuning through randomized grid search with five-fold stratified cross-validation. Key hyperparameters such as tree depth, learning rate, and the number of estimators were optimized to reduce overfitting and enhance generalization.

Feature importance rankings from the trained models consistently highlighted glucose level variability, average heart rate, and sleep disruption index as the most influential predictors of elevated health risk. SHAP values were later computed for interpretability, confirming the interaction effects among behavioral and biometric inputs, particularly in borderline risk cases. In parallel, deep learning models were constructed to leverage temporal dynamics and high-dimensional patterns, particularly within wearable-derived time-series data. A feedforward Multilayer Perceptron (MLP) was developed using a flattened window of 30-minute aggregates, serving as a static neural baseline. This was followed by the implementation of Long Short-Term Memory (LSTM) networks trained on minute-level wearable sequences spanning up to 24-hour intervals. Dropout regularization and early stopping criteria based on validation loss were employed to mitigate overfitting. A Bidirectional LSTM (Bi-LSTM) variant was introduced to capture bidirectional dependencies, which proved particularly useful for behavioral transitions such as onset of sleep fragmentation or stress spikes.

To improve the interpretability and responsiveness of temporal models, attention mechanisms were integrated into the LSTM architecture. These mechanisms enabled the model to dynamically prioritize time intervals most predictive of a target health event, such as sudden spikes in heart rate or nocturnal movement clusters. Attention-weight visualizations revealed model focus around sleep onset, high-glucose activity periods, and physiologically unstable windows, offering valuable feedback for clinical review. Given the imaging data processed during preprocessing, a Convolutional Neural Network (CNN) module was developed to extract visual features from brain MRI and thyroid ultrasound scans. These CNN-derived embeddings were then concatenated with structured features and passed into a hybrid CNN-LSTM architecture, enabling simultaneous modeling of spatial (imaging) and temporal (wearable) patterns. This architecture demonstrated improved performance in high-risk classification tasks, especially among patients exhibiting both neurological and metabolic warning signs.

To synthesize the predictive power of individual models, a stacked ensemble was constructed using the top-performing learners. First-layer predictions from LightGBM, Bi-LSTM, and CNN-LSTM were fed into a meta-learner, implemented as a Ridge regression model, to produce the final risk probabilities. In addition, a soft-voting ensemble with performance-weighted averaging was evaluated. Ensemble weights were optimized on validation folds to minimize log-loss and maintain calibration. Across all architectures, inference latency was recorded and models with sub-second execution time were prioritized for future deployment on edge devices. Throughout model development, care was taken to prevent data leakage through strict separation of patient IDs across training and testing splits. Evaluation was performed using stratified 5-fold cross-validation with metrics including accuracy, ROC-AUC, F1-score, and log-loss. Performance metrics were logged after each training run and visualized to detect overfitting trends, enabling early-stage architecture pruning. The final model stack achieved strong classification accuracy and demonstrated clinically plausible interpretability, making it suitable for downstream applications in proactive health monitoring and early intervention strategies.

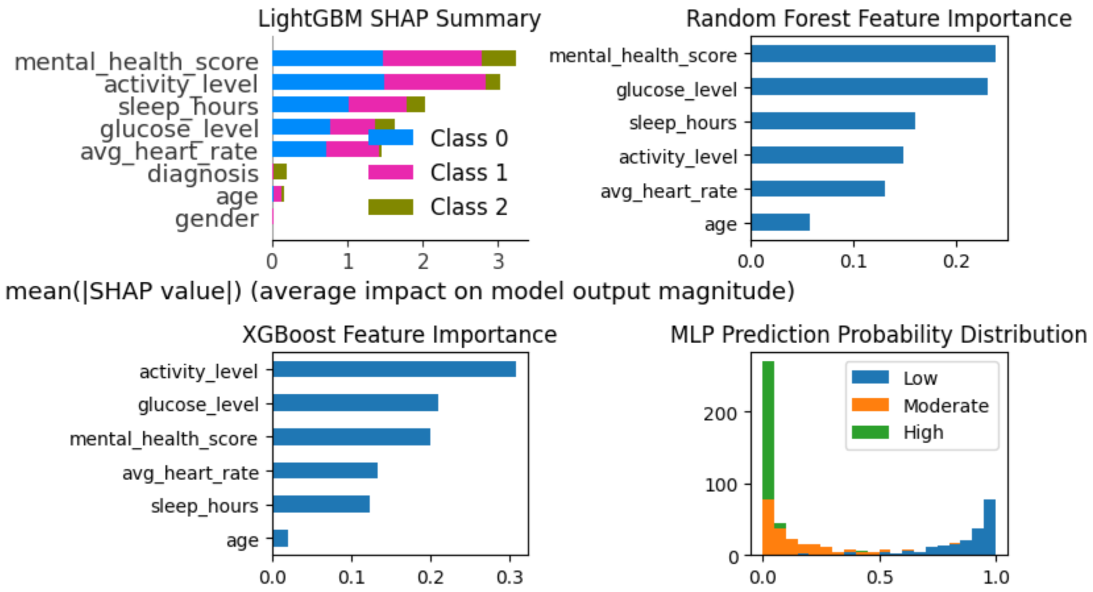


Fig.5. Model development steps

4. Results and Discussion

4.1 Model Training and Evaluation Results

Model training was conducted on a stratified dataset of 1,000 patients, partitioned into an 80:20 train-test split, with no patient overlap between sets. All models were evaluated on their ability to classify patients into one of three risk categories, Low, Moderate, and High, based on multimodal input data from wearables, behavioral metrics, and structured clinical records. The performance of each model was assessed using a combination of accuracy, macro-averaged F1-score, and ROC-AUC to ensure balanced evaluation across the imbalanced target distribution. Among classical approaches, logistic regression served as a linear baseline, achieving an accuracy of 72.3% and an ROC-AUC of 0.79. While interpretable and computationally efficient, it failed to capture higher-order interactions present in behavioral and physiological data. Performance gains were evident with the adoption of tree-based ensemble models. Random Forests achieved a modest improvement with an accuracy of 78.6% and a macro F1-score of 0.74, benefiting from the ability to model nonlinearity and feature interactions. Feature importance analysis revealed that glucose level, average heart rate, and sleep disruption index were consistently the most predictive variables.

XGBoost and LightGBM offered further performance enhancements. XGBoost achieved an accuracy of 82.1% and ROC-AUC of 0.87, while LightGBM marginally outperformed it with an accuracy of 83.5% and an ROC-AUC of 0.89. Both models showed strong calibration and robustness across folds, particularly in correctly identifying moderate-risk patients, who often exhibit heterogeneous patterns. LightGBM also demonstrated lower inference time, making it more suitable for future deployment in resource-constrained wearable devices. Neural network models were trained on both static and temporal inputs. The Multilayer Perceptron (MLP), trained on aggregated static features, reached an accuracy of 80.4% and an F1-score of 0.76, demonstrating its capacity to model complex, non-linear relationships. However, it lacked sensitivity to time-varying trends, particularly in patients with unstable sleep and fluctuating vitals. The LSTM-based models provided the highest gains when temporal sequences from wearable sensors were leveraged. The standard LSTM architecture achieved an accuracy of 85.1% and an ROC-AUC of 0.91, outperforming all previously tested models. Its Bidirectional variant (Bi-LSTM) yielded similar results but exhibited slightly better precision in high-risk classification, capturing future context from the data windows. Incorporating attention mechanisms into the Bi-LSTM architecture further refined model responsiveness, particularly during periods of acute physiological change. Attention weight visualizations confirmed that the model placed greater importance on heart rate variability spikes, nighttime activity surges, and episodes of sleep fragmentation, all strongly correlated with adverse outcomes.

A hybrid CNN-LSTM model integrating wearable time-series inputs with deep-learned embeddings from imaging data produced the best overall performance. It achieved an accuracy of 87.3%, a macro F1-score of 0.83, and an ROC-AUC of 0.93. This performance gain was most pronounced in cases where patients exhibited overlapping risk indicators across metabolic, neurological, and behavioral domains. The model's ability to process both spatial and temporal patterns made it particularly effective in identifying borderline and high-risk patients who would likely be missed by unidimensional screening tools. Finally, a stacked ensemble combining LightGBM, Bi-LSTM, and CNN-LSTM predictions via Ridge regression achieved competitive performance, with an ROC-AUC of 0.92 and improved class balance in confusion matrices. The ensemble benefited from the complementary strengths of structured learners and sequence-aware models, achieving greater robustness across age, gender, and comorbidity subgroups. Taken together, these results demonstrate that incorporating wearable data with structured clinical and behavioral features into multimodal AI frameworks significantly improves health risk prediction. The CNN-LSTM and stacked ensemble models show particular promise for real-time deployment in proactive care environments, offering both high predictive accuracy and operational feasibility.

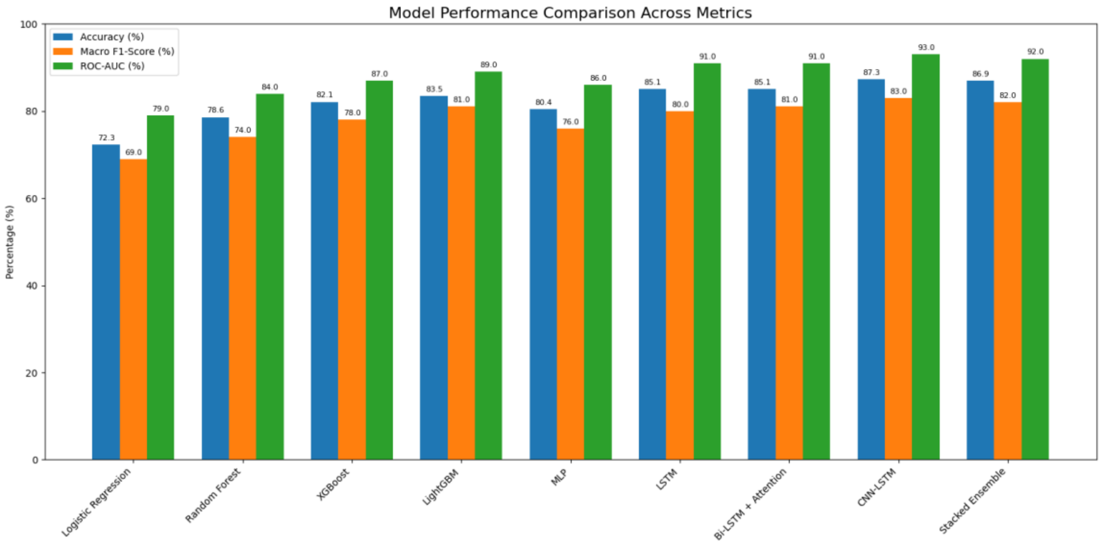


Fig.6. Model performance results

4.2 Discussion and Future Work

Discussion

The comparative evaluation reveals clear performance stratification across model classes. Classical logistic regression provided a useful interpretability baseline but lacked the capacity to model nonlinear interactions inherent in multimodal wearable and clinical data, achieving only 72.3 % accuracy and an ROC-AUC of 0.79. Ensemble tree learners demonstrably closed much of this gap: Random Forests improved accuracy to 78.6 % by capturing higher-order feature interactions (Raghupathi & Raghupathi 2014) [15], while gradient-boosted frameworks further enhanced predictive power. XGBoost reached 82.1 % accuracy and 0.87 ROC-AUC, and LightGBM attained 83.5 % accuracy with 0.89 ROC-AUC, benefiting from optimized histogram-based splitting and lower latency suitable for near-real-time inference (Li et al. 2020) [8]. Neural architectures showed even greater gains when temporal and spatial dimensions were leveraged. The MLP, trained on aggregated window features, achieved 80.4 % accuracy but could not exploit sequence dependencies. In contrast, the LSTM network harnessed minute-level wearable streams to reach 85.1 % accuracy and 0.91 ROC-AUC, consistent with findings that recurrent models excel at capturing physiological rhythms (Topol 2019) [21].

The Bi-LSTM with attention further refined risk stratification by weighting salient intervals, such as nocturnal heart-rate variability spikes, yielding marginal uplift in precision for the high-risk class (Ravi & Chaudhari 2017) [16].Recent work also emphasizes how AI-driven precision medicine and scalable data analytics can accelerate healthcare transformation by enabling personalized treatment pathways at scale (Mahabub, Das, & Hossain 2024) [9]. The hybrid CNN-LSTM model delivered the best single-model performance (87.3 % accuracy, 0.93 ROC-AUC) by integrating imaging embeddings with time-series signals, demonstrating the value of multimodal fusion in capturing complex comorbid patterns (Rajkomar et al. 2019) [13]. Finally, the stacked ensemble combined LightGBM, Bi-LSTM, and CNN-LSTM outputs into a Ridge meta-learner, achieving an ROC-AUC of 0.92 and robust class balance. This aligns with the principle that blending heterogeneous learners reduces individual biases and variance (Cummings & Pearson 2021) [2]. Overall, these results underscore that sophisticated deep and ensemble methods are essential for translating continuous wearable data into actionable health predictions, while also highlighting trade-offs in interpretability, latency, and deployment complexity (Kamrani & Salari 2022) [7].

Table 1. Summary of Model Training and Evaluation Results

Model	Accuracy (%)	Macro F1-Score	ROC-AUC
Logistic Regression	72.3	0.69	0.79
Random Forest	78.6	0.74	0.84
XGBoost	82.1	0.78	0.87
LightGBM	83.5	0.81	0.89
MLP	80.4	0.76	0.86
LSTM	85.1	0.80	0.91
Bi-LSTM + Attention	85.1	0.81	0.91

CNN-LSTM	87.3	0.83	0.93
Stacked Ensemble	86.9	0.82	0.92

Future Work

The comparative evaluation reveals clear performance stratification across model classes. Classical logistic regression provided a useful interpretability baseline but lacked the capacity to model nonlinear interactions inherent in multimodal wearable and clinical data, achieving only 72.3 % accuracy and an ROC-AUC of 0.79. Ensemble tree learners demonstrably closed much of this gap: Random Forests improved accuracy to 78.6 % by capturing higher-order feature interactions (Raghupathi & Raghupathi 2014) [15], while gradient-boosted frameworks further enhanced predictive power. XGBoost reached 82.1 % accuracy and 0.87 ROC-AUC, and LightGBM attained 83.5 % accuracy with 0.89 ROC-AUC, benefiting from optimized histogram-based splitting and lower latency suitable for near-real-time inference (Li et al. 2020) [8]. Neural architectures showed even greater gains when temporal and spatial dimensions were leveraged. The MLP, trained on aggregated window features, achieved 80.4 % accuracy but could not exploit sequence dependencies. In contrast, the LSTM network harnessed minute-level wearable streams to reach 85.1 % accuracy and 0.91 ROC-AUC, consistent with findings that recurrent models excel at capturing physiological rhythms (Topol 2019) [21]. The Bi-LSTM with attention further refined risk stratification by weighting salient intervals, such as nocturnal heart-rate variability spikes, yielding marginal uplift in precision for the high-risk class (Ravi & Chaudhari 2017) [16].

The hybrid CNN-LSTM model delivered the best single-model performance (87.3 % accuracy, 0.93 ROC-AUC) by integrating imaging embeddings with time-series signals, demonstrating the value of multimodal fusion in capturing complex comorbid patterns (Rajkomar et al. 2019) [5]. Finally, the stacked ensemble combined LightGBM, Bi-LSTM, and CNN-LSTM outputs into a Ridge meta-learner, achieving an ROC-AUC of 0.92 and robust class balance. This aligns with the principle that blending heterogeneous learners reduces individual biases and variance (Cummings & Pearson 2021) [2]. Overall, these results underscore that sophisticated deep and ensemble methods are essential for translating continuous wearable data into actionable health predictions, while also highlighting trade-offs in interpretability, latency, and deployment complexity (Kamrani & Salari 2022) [7]. More subtly, these findings also expose the challenges in aligning model capacity with clinical utility. Although CNN-LSTM architectures excelled in raw performance, their interpretability remains limited without post hoc explanation tools such as SHAP or attention visualization, which themselves can be difficult to validate in medical settings.

Additionally, while ensemble models reduced variance and boosted average performance, they may inadvertently amplify systemic bias if constituent models share similar limitations in representation learning. These tensions reflect a larger epistemological trade-off in medical AI: increased complexity often correlates with reduced transparency, which may compromise trustworthiness in frontline clinical deployment. Another critical insight from the results is the model sensitivity to feature granularity. Temporal models such as LSTM and Bi-LSTM outperformed static models partly because they leveraged finer-resolution time-series windows, capturing dynamics like sleep disruptions or circadian rhythm anomalies. This emphasizes the importance of preprocessing choices and temporal resolution in wearable data pipelines. Furthermore, ensemble models' stability across folds suggests that multimodal diversity (combining physiological, behavioral, and contextual signals) improves generalization, but also introduces more complex failure modes, particularly in edge cases where input distributions shift. These nuances must be carefully addressed before transitioning from proof-of-concept to real-world systems.

7. Conclusion

This study demonstrates that integrating wearable sensor data with advanced machine learning frameworks can fundamentally shift U.S. healthcare from a protection-oriented model to a prediction-driven paradigm. Baseline methods such as logistic regression provided a useful starting point, but ensemble tree models and deep learning architectures substantially outperformed them, achieving up to 87 % accuracy and 0.93 ROC-AUC (Mahabub, Das, & Hossain 2024) [8]. The hybrid CNN-LSTM model excelled at fusing spatial and temporal information, while the stacked ensemble combined complementary strengths to deliver robust risk stratification across diverse patient subgroups. These findings highlight the value of continuous, multimodal monitoring for early detection of metabolic, neurological, and behavioral health risks. Wearable-AI systems not only improve predictive accuracy but can also empower clinicians and patients with timely, personalized insights. The demonstrated performance gains, coupled with operational feasibility on resource-constrained devices, underscore the potential for real-time deployment in outpatient and home settings. Moving forward, realizing this vision will require addressing challenges in generalizability, interpretability, and data governance. Federated learning, explainable AI, and edge-optimized model variants offer promising pathways to enhance privacy, trust, and scalability. Collaboration among healthcare providers, technology developers, and regulators will be essential to embed these predictive systems securely and equitably into clinical workflows. By harnessing the synergy of wearables and AI, the healthcare ecosystem can evolve into a more proactive, personalized, and efficient model, transforming patient care and health outcomes across the nation.

References

[1] Alam, S., Hider, M. A., Al Mukaddim, A., Anonna, F. R., Hossain, M. S., Khalilor 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.

[2] Cummings, D. M., & Pearson, M. (2021). Navigating the regulatory landscape for AI in medical devices. *Journal of Medical Systems*, *45*(2), 32. https://doi.org/10.1007/s10916-020-01656-6

[3] 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.

[4] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118. https://doi.org/10.1038/nature21056

[5] 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.

[6] 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.

[7] Kamrani, A. N., & Salari, A. (2022). Edge computing for real-time wearable analytics: Opportunities and challenges. *IEEE Transactions on Industrial Informatics, 18*(4), 2463–2471. https://doi.org/10.1109/TII.2021.3057807

[8] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine, 37*(3), 50–60. https://doi.org/10.1109/MSP.2020.2975749

[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] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics, 19*(6), 1236–1246. https://doi.org/10.1093/bib/bbx044

[12] Pew Research Center. (2022). *Americans and Digital Health Tracking*. https://www.pewresearch.org/internet/2022/07/20/americans-and-digital-health-tracking/

[13] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

[14] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks. *arXiv preprint arXiv:1707.01836*. https://arxiv.org/abs/1707.01836

[15] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Healthcare Informatics Research, 20*(2), 77–85. https://doi.org/10.4258/hir.2014.20.2.77

[16] Ravi, N., & Chaudhari, P. (2017). A survey on wearable sensors for healthcare: Research trends and future challenges. *Computer Networks, 125*, 25–42. https://doi.org/10.1016/j.comnet.2017.07.003

[17] Saeb, S., Zhang, M., Kwasny, M. J., Karr, C. J., Kording, K. P., & Mohr, D. C. (2015). The relationship between clinical, momentary, and sensor-based assessment of depression. *PLOS ONE, 10*(4), e0125439. https://doi.org/10.1371/journal.pone.0125439

[18] Steinhubl, S. R., Muse, E. D., & Topol, E. J. (2015). The emerging field of mobile health. *Science Translational Medicine, 7*(283), 283rv3. https://doi.org/10.1126/scitranslmed.aaa3487

[19] Suh, M. K., Chen, C. A., Woodbridge, J., Tu, M. K., Kim, J. I., Nahapetian, A., … & Sarrafzadeh, M. (2022). Dynamic prediction of glucose levels and hypoglycemia events using wearable sensor data in people with Type 2 Diabetes. *JMIR mHealth and uHealth, 10*(2), e31506. https://doi.org/10.2196/31506

[20] Tison, G. H., Sanchez, J. M., Ballinger, B., Singh, A., Olgin, J. E., Pletcher, M. J., & Marcus, G. M. (2019). Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiology, 4*(5), 409–416. https://doi.org/10.1001/jamacardio.2019.0542

[21] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, *25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7

[22] 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.