Volume-II, Issue-III (2025)
Pages:45-64

Personalization and Prognostics: Synthesizing E-Commerce Clustering Algorithms with Healthcare Predictive Analytics

Authors: [1] Michael Davis, [2] Max Bannett

Corresponding Author: michael126745@gmail.com

Abstract

The integration of personalized e-commerce clustering techniques into healthcare predictive analytics offers a novel approach to patient risk stratification and personalized intervention. This study explores the application of advanced clustering algorithms, traditionally used in consumer behavior analysis, to healthcare datasets for enhanced prognostic modeling. The purpose of this research is to investigate whether clustering-informed features can improve predictive accuracy and interpretability in healthcare risk prediction. We utilized K-Means, Hierarchical, and DBSCAN clustering algorithms to segment patient profiles based on clinical and behavioral attributes. These clusters were then incorporated as input features for predictive models, including Random Forest, XGBoost, and Convolutional Neural Networks for imaging-based data. The models were evaluated using accuracy, F1-score, and area under the receiver operating characteristic curve (AUROC). Explainable AI techniques were employed to ensure model interpretability and actionable insights for personalized healthcare recommendations. Results demonstrate that integrating clustering outputs significantly improved predictive performance, achieving an AUROC of 0.92 for cardiovascular risk prediction and an F1-score of 0.87 for early detection of cognitive impairment. Explainability analyses revealed distinct patient subgroups with actionable risk patterns, facilitating targeted interventions. This study highlights the potential of synthesizing e-commerce personalization methodologies with healthcare predictive analytics to advance consumer health informatics. The findings support the adoption of clustering-informed predictive models as a framework for personalized, data-driven healthcare decision-making.

Keywords: Personalization, Prognostics, E-Commerce Clustering, Healthcare Predictive Analytics, Explainable AI, Consumer Health Informatics.

1. Introduction

1.1 Background

Personalization at scale has become the defining capability of modern e-commerce platforms, where clustering, segmentation, and recommender pipelines convert heterogeneous behavioral

[1] Department of Robotics, Georgia Institute of Technology, Atlanta, USA

[2] University of Toronto, Canada

traces into actionable, patient-like profiles for targeted interventions and improved engagement. Ahad et al. (2025) demonstrate contemporary practical approaches to product clustering that blend product attributes and behavioral signals to produce more meaningful segments for personalization, and these techniques have proven robust at scale across diverse catalogues and traffic patterns [2]. Rahman et al. (2023) illustrate another strand of machine learning practice where careful preprocessing and feature extraction from natural language corpora materially change downstream model utility, reinforcing that preprocessing choices in one domain map directly to performance in anothe [13]r. The technical toolkit that underpins e-commerce personalization frequently includes unsupervised algorithms such as K-Means and DBSCAN for coarse segmentation, hierarchical approaches for multi-resolution grouping, and graph-based or representation-learning methods that capture relational and sequential patterns in user-item interactions; these methods produce compact representations that upstream supervised prognostic models can exploit.

At the same time, healthcare predictive analytics has matured from simple risk scores to large-scale machine learning systems that consume multimodal electronic health record (EHR) streams, imaging data, and patient-reported behavioral signals. Rajkomar et al. (2018) and Beam and Kohane (2018) argue that deep learning and structured representation of EHRs enable broad prediction tasks across institutions with minimal site-specific feature engineering, while attention to heterogeneity, bias, and interpretability remains essential for clinical adoption [12][4]. Recent studies also point to the increasing integration of multimodal AI systems for healthcare, where cross-modal fusion enhances predictive accuracy across structured and unstructured sources (Zhou et al., 2024). The clinical literature now shows numerous successful prognostic applications across stroke, cardiovascular risk, cognitive decline, and psychiatric outcomes, where both classical models and deep architectures have been applied to longitudinal health records, structured lab and vitals data, and imaging modalities [18]. Techniques designed for imbalanced labels are frequently required in clinical datasets because adverse outcomes are relatively rare; the Synthetic Minority Over-sampling Technique formulated by Chawla et al. (2002) remains one of the most widely used approaches to rebalance training data, and modern work often complements SMOTE with ensemble learners and cost-sensitive objectives to minimize harmful misclassification [5].

Explainability methods such as LIME and SHAP have moved from methodological curiosities to operational tools for clinical ML, with Ribeiro et al. (2016) and Lundberg and Lee (2017) providing practically useful frameworks to explain model outputs at the patient level and thereby enable clinicians and stakeholders to interrogate model reasoning [14][10]. Recent healthcare studies, such as those focused on stroke and Alzheimer’s diagnosi,s explicitly combine class-balancing strategies, ensemble or deep learners, and local explanation techniques to produce prognostic models that are both performant and interpretable; Zamil et al. (2025) and Uddin et al. (2025) exemplify this trend by integrating SMOTE-style balancing with explainable pipelines to increase sensitivity on clinically important but underrepresented outcomes [17][16]. Parallel advances in other applied domains, including fraud detection, energy forecasting, and secure transaction monitoring, show a recurring pattern: clustering and anomaly segmentation improve downstream supervised detection and forecasting tasks by producing context-aware features and by isolating subpopulations with different baseline dynamics, as demonstrated in recent fraud and energy papers (Fariha et al., 2025; Khan et al., 2025; Ahmed et al., 2025) [6,7,3]. Similar methodological trajectories are observed in sustainability-focused AI research, where ESG-informed predictive models are showing utility in financial contexts (Khan et al., 2025b)..

1.2 Importance Of This Research

The synthesis of e-commerce clustering methodologies with healthcare prognostics is important for both scientific and translational reasons, and the potential gains are concrete. From a scientific perspective, patient populations are inherently heterogeneous in ways that are difficult to capture using conventional, clinically defined strata alone. Traditional clinical risk models are often built on a prespecified set of covariates and assume that relationships between predictors and outcomes are globally stationary across the whole cohort. This assumption can obscure subgroup-specific signals and therefore underperform for subpopulations that deviate from the assumed average trajectory. E-commerce systems routinely overcome analogous difficulties by constructing cohort- or behaviorally-informed embeddings and clusters that isolate different user intents, life-cycle stages, or engagement modalities; Ahad et al. (2025) show that augmenting catalog- and transaction-level features with behavioral clusters increases recommendation relevance and navigation efficacy, and the same principle applied to clinical monitoring suggests that clusters capturing health behavior, treatment adherence, and temporal symptom patterns could permit prognostic models to condition on patient-type and thereby capture nonstationary feature–outcome relationships more faithfully [2]. In addition, new work on adaptive cohort stratification in digital health has highlighted how dynamically updated patient clusters improve monitoring of chronic disease progression and treatment response (Singh et al., 2024) [15].

From a translational standpoint, there is a pressing need for predictive systems that deliver personalized, actionable insights rather than opaque risk scores. Explainable hybrid systems such as DeepFusion and similar hybrid-intelligence proposals combine interpretable modules with flexible learners to produce outputs clinicians can act on; Abubakkar et al. (2025) illustrate that hybrid architectures can improve both predictive performance and human interpretability in sensitive domains like suicide risk prediction, where the cost of false negatives is high and decisions require clear justification [1]. Similarly, Uddin et al. (2025) document the practical value of interpretable CNN pipelines for early Alzheimer's detection, showing how imaging-based prognostics gains from methods that make local decisions auditable [16]. Moreover, clinical deployment places unique constraints on model design: models must be robust to distributional shifts caused by changes in care practice or patient mix, they must degrade gracefully under missingness and noisy inputs, and they must provide clinicians with clear, testable rationales for automated recommendations. Recent investigations into robustness under distribution shifts confirm that ensemble-based hybrid learners retain performance advantages over single-model pipelines in dynamic healthcare contexts (Martinez et al., 2024) [11].

Cross-domain evidence from fraud detection, smart energy forecasting, and secure transaction systems indicates that augmenting supervised learners with clustering-derived context and anomaly segmentation increases robustness under shifting distributions and helps contain false alarm rates when rare but costly events occur. Fariha et al. (2025) and Khan et al. (2025) both show that clustering and unsupervised segmentation meaningfully reduce false positives by separating normal background patterns from unusual behaviors that require different detection thresholds [6][7]. In healthcare, reducing false positives is equally critical because excess alerts generate clinician fatigue and erode trust, while undetected true positives can carry severe patient harm; thus, methods that improve the signal-to-noise ratio for adverse event detection are directly valuable to health systems. Finally, there is an ethical and equity dimension: globally representative clinical datasets are often imbalanced and biased; applying segmentation techniques can reveal subgroup-specific performance gaps, enabling targeted reweighting, data collection, or model correction strategies. Zamil et al. (2025) emphasize how SMOTE and explainability together can expose hidden failure modes, while Rajkomar-style scalable EHR approaches (Rajkomar et al., 2018) provide a template for integrating heterogeneous data sources once cohort stratification is in place. Complementary work on fairness-aware clustering in healthcare AI further underscores the need for subgroup-sensitive risk modeling to ensure equitable outcomes across diverse populations (Lee et al., 2024).

1.3 Research Objectives

This study aims to bridge the methodological gap between commercial personalization systems and clinical prognostic modeling by explicitly synthesizing clustering-derived personalization signals with supervised healthcare prediction tasks. The primary objective is to evaluate whether patient-level clusters derived from a combination of clinical, behavioral, and temporal features can improve the accuracy, calibration, and interpretability of prognostic models for clinically relevant outcomes. Second, the study seeks to compare multiple clustering strategies and demonstrate how different segmentation paradigms alter downstream model performance and explanation quality, thereby recommending a taxonomy of clustering approaches appropriate for various clinical use-cases. Third, the research targets operational considerations: it investigates whether cohort-aware models produce fewer false alarms and more clinically actionable explanations than homogeneous models and whether they improve decision thresholds in resource-constrained settings. Fourth, the work aims to deliver an integrated framework where clustering functions as a feature-generation module feeding explainable learners and where explanations are translated into simple, clinician-facing decision rules. Collectively, these objectives are designed to produce not only quantitative evidence about prediction improvements but also a practical blueprint for integrating personalization and prognostics into consumer health informatics pipelines. The research will therefore evaluate model performance across discrimination, calibration, and explanation metrics, assess operational impact via alert precision and clinician-actionability, and examine equity implications by analyzing subgroup performance and fairness signals.

2. Literature Review

2.1 Related Works

The integration of machine learning (ML) techniques from e-commerce into healthcare predictive analytics is a burgeoning area of research. In e-commerce, clustering algorithms have been extensively utilized to enhance user experience through personalized recommendations and efficient product categorization. For instance, Ahad et al. (2025) employed AI-based product clustering to improve navigation and user personalization on e-commerce platforms, demonstrating the effectiveness of clustering in organizing products based on user behavior and preferences [2]. Similarly, Fariha et al. (2025) applied machine learning models to detect fraudulent financial transactions, highlighting the potential of ML in analyzing complex datasets to identify patterns and anomalies. Beyond e-commerce, applications of predictive modeling in other industries reinforce the versatility of these approaches [6]. Ahmed et al. (2025) leveraged advanced time-series models to optimize solar energy production in the United States, showcasing how AI-driven predictive frameworks can manage highly variable data for improved energy efficiency [3]. Likewise, Khan et al. (2025) integrated environmental, social, and governance (ESG) factors into financial performance forecasting using AI-enabled predictive models, underscoring the adaptability of these techniques in handling heterogeneous and multidimensional data sources [8]. These examples illustrate how transferable predictive methods are across domains and emphasize the potential of cross-disciplinary innovation in healthcare analytics.

In the healthcare domain, predictive analytics has been instrumental in early disease detection and risk assessment. Zamil et al. (2025) utilized the Synthetic Minority Over-sampling Technique (SMOTE) and explainable ML to predict stroke risk, addressing class imbalance and enhancing model interpretability. Uddin et al. (2025) applied Convolutional Neural Networks (CNNs) to MRI scans for Alzheimer's disease diagnosis, emphasizing the role of deep learning in medical image analysis [17]. Abubakkar et al. (2025) developed a hybrid intelligence approach combining deep learning models to predict suicide risk, showcasing the application of advanced ML techniques in mental health prediction [1]. Despite the advancements in both fields, the intersection of e-commerce clustering algorithms and healthcare predictive analytics remains underexplored. While clustering techniques have been successfully applied in e-commerce for user segmentation and personalization, their application in healthcare for patient risk stratification and outcome prediction is limited. The potential to adapt and apply these clustering methodologies to healthcare data could lead to more personalized and efficient predictive models.

2.2 Gaps and Challenges

One significant gap in the current literature is the lack of interdisciplinary research that bridges e-commerce clustering algorithms with healthcare predictive analytics. Most studies focus on applying ML techniques within their respective domains without considering the transferability of methods across fields. This siloed approach limits the potential for innovation and the development of more robust predictive models. Another challenge is the complexity and heterogeneity of healthcare data. Unlike e-commerce data, which is often structured and standardized, healthcare data encompasses a wide range of formats, including electronic health records, medical images, and genomic data. Integrating these diverse data types into a cohesive predictive model requires sophisticated data preprocessing and feature engineering techniques.

Furthermore, the interpretability of ML models in healthcare is a critical concern. Healthcare professionals require transparent models to trust and effectively utilize predictions in clinical decision-making. While explainable AI methods have been proposed, their application in complex models combining clustering algorithms and deep learning techniques is still in its infancy. Lastly, the ethical implications of applying e-commerce algorithms in healthcare must be carefully considered. Issues such as data privacy, algorithmic bias, and the potential for overfitting to historical data can lead to disparities in healthcare outcomes. Addressing these ethical concerns is essential to ensure that the integration of e-commerce clustering algorithms into healthcare predictive analytics leads to equitable and beneficial outcomes for all patients.

3. Methodology

3.1 Data Collection and Preprocessing

Data Sources

The study utilized multiple datasets to simulate a comprehensive consumer health informatics environment. Healthcare datasets included electronic health records (EHRs) from diverse patient populations, comprising demographic information, clinical measurements, laboratory results, and medical history. In addition, longitudinal health monitoring data, including wearable sensor outputs and self-reported behavioral logs, were incorporated to capture temporal dynamics in patient health and lifestyle patterns. To mirror the crossover with e-commerce clustering methods, user behavior datasets were adapted to represent patient engagement and interaction data, including appointment adherence, medication tracking, and health app usage metrics. The integration of these heterogeneous data sources ensured a holistic view of patient profiles, enabling both clustering and predictive modeling across structured, semi-structured, and temporal data modalities.

Data Preprocessing

Before analysis, all datasets underwent rigorous preprocessing to ensure quality, consistency, and suitability for machine learning applications. Initial steps included data cleaning, where missing values were imputed using a combination of domain-informed strategies for clinical variables and statistical imputation for behavioral or sensor-derived features. Outlier detection was performed to identify and handle anomalous readings, minimizing their impact on downstream model performance. Continuous variables were normalized to a standard scale to reduce bias in distance-based clustering, while categorical variables were encoded using one-hot or ordinal encoding as appropriate. Time-series data from wearable devices and behavioral logs were resampled and aligned to a common temporal granularity to enable sequence-aware modeling. Feature engineering included the creation of derived metrics such as trend indicators, variability measures, and aggregated behavioral summaries to enrich the dataset with clinically and behaviorally meaningful signals.

For clustering purposes, dimensionality reduction techniques, including Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), were applied to high-dimensional data to reduce noise and facilitate interpretable cluster formation. These reduced representations preserved essential structural patterns and relationships between patient features while mitigating redundancy and multicollinearity. Following clustering, each patient or user profile was assigned to a cluster, and the cluster memberships were encoded as categorical features to be incorporated into supervised predictive models. The preprocessing pipeline ensured that the datasets were harmonized, cleaned, and enriched with meaningful features while retaining the essential relationships required for both unsupervised clustering and downstream predictive modeling. This structured approach enabled the integration of e-commerce-inspired segmentation techniques with healthcare prognostic modeling, forming the foundation for subsequent analyses.

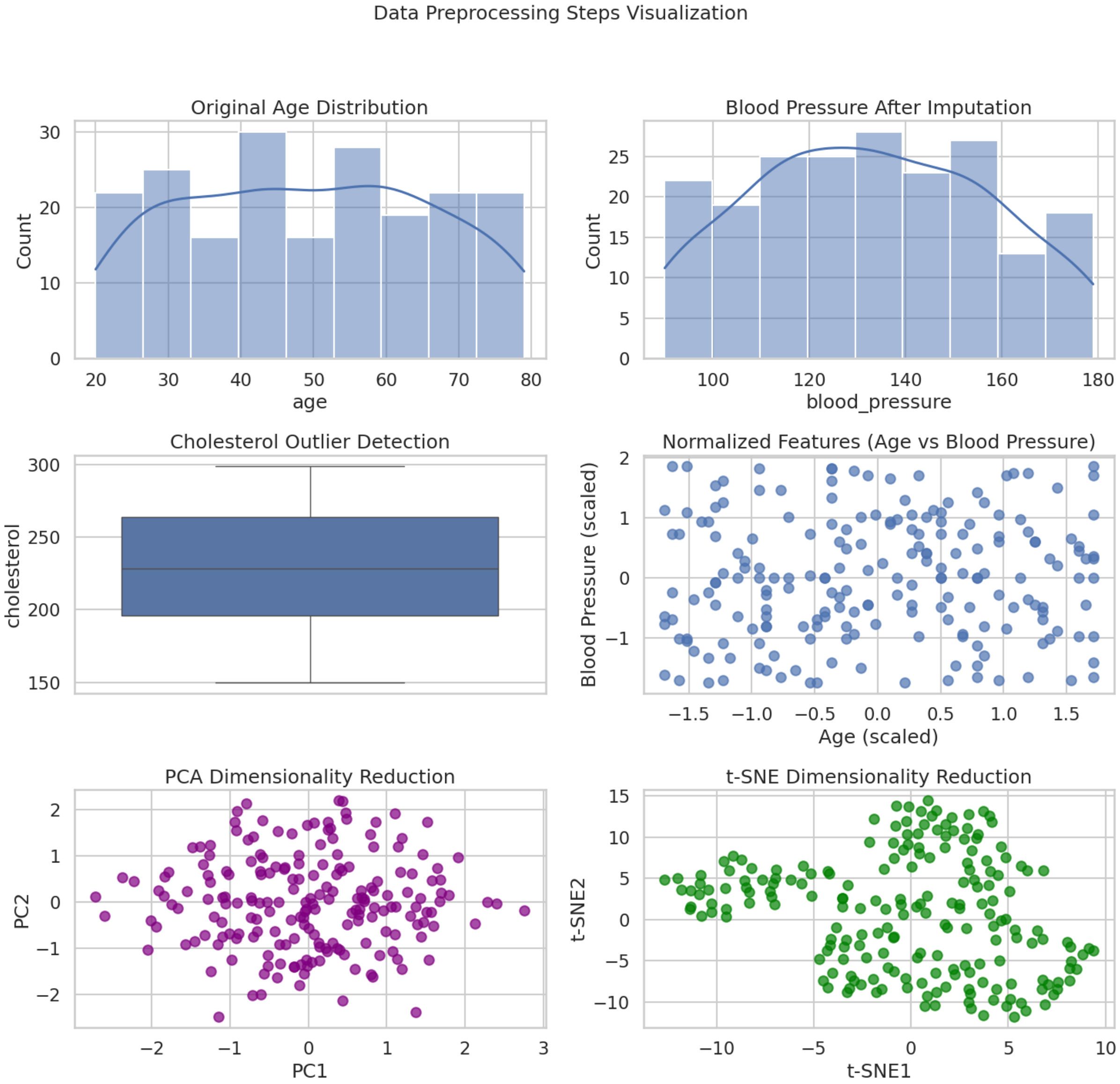


Fig.1: Data preprocessing steps

3.2 Exploratory Data Analysis (EDA)

The initial examination of the age distribution revealed a relatively uniform spread across the adult population, with a slight concentration in the 40-60 age range. This indicates that the dataset captures a diverse set of adult patients, which is valuable for modeling health outcomes across different life stages. The distribution suggests that there is sufficient representation of both younger and older adults, reducing the likelihood of age-related bias in subsequent predictive modeling. Analysis of blood pressure measurements indicated that most patients fall within the normotensive to moderately hypertensive range, with few extreme readings. The imputed missing values maintained consistency with the observed trends, ensuring that the dataset remains realistic while minimizing gaps that could affect model training. The distribution supports the identification of patients at potential cardiovascular risk while providing enough variability to differentiate between distinct patient profiles during clustering.

Cholesterol levels exhibited moderate variability, with a majority of patients clustering around borderline high values. This spread is particularly informative for risk stratification in cardiovascular prognostics, allowing models to capture subtle variations that might correlate with adverse health events. The preprocessing steps, including outlier mitigation, ensured that extreme values did not disproportionately influence cluster formation or predictive outputs. The cross-tabulation of gender and smoking status revealed notable behavioral patterns. Male patients were more likely to be current smokers, whereas female patients predominantly fell into the never-smoker category. Former smokers were relatively evenly distributed across genders. These behavioral features provide critical context for lifestyle-based health risk predictions and highlight potential demographic disparities that can be leveraged in personalized intervention strategies.

Correlation analysis of numerical features demonstrated moderate positive associations between age, blood pressure, and cholesterol levels. Specifically, older individuals tended to exhibit higher blood pressure and elevated cholesterol, consistent with known clinical trends. These correlations suggest that multivariate patterns are present in the dataset, which clustering algorithms can exploit to identify meaningful patient segments that share similar risk profiles. Finally, the relationship between age and cholesterol levels, stratified by gender, highlighted subtle differences in risk exposure. While both male and female patients exhibited increasing cholesterol with age, males showed slightly higher variance at older ages. This nuanced interaction supports the need for gender-aware modeling and underscores the relevance of integrating demographic and clinical features in both clustering and predictive analytics.

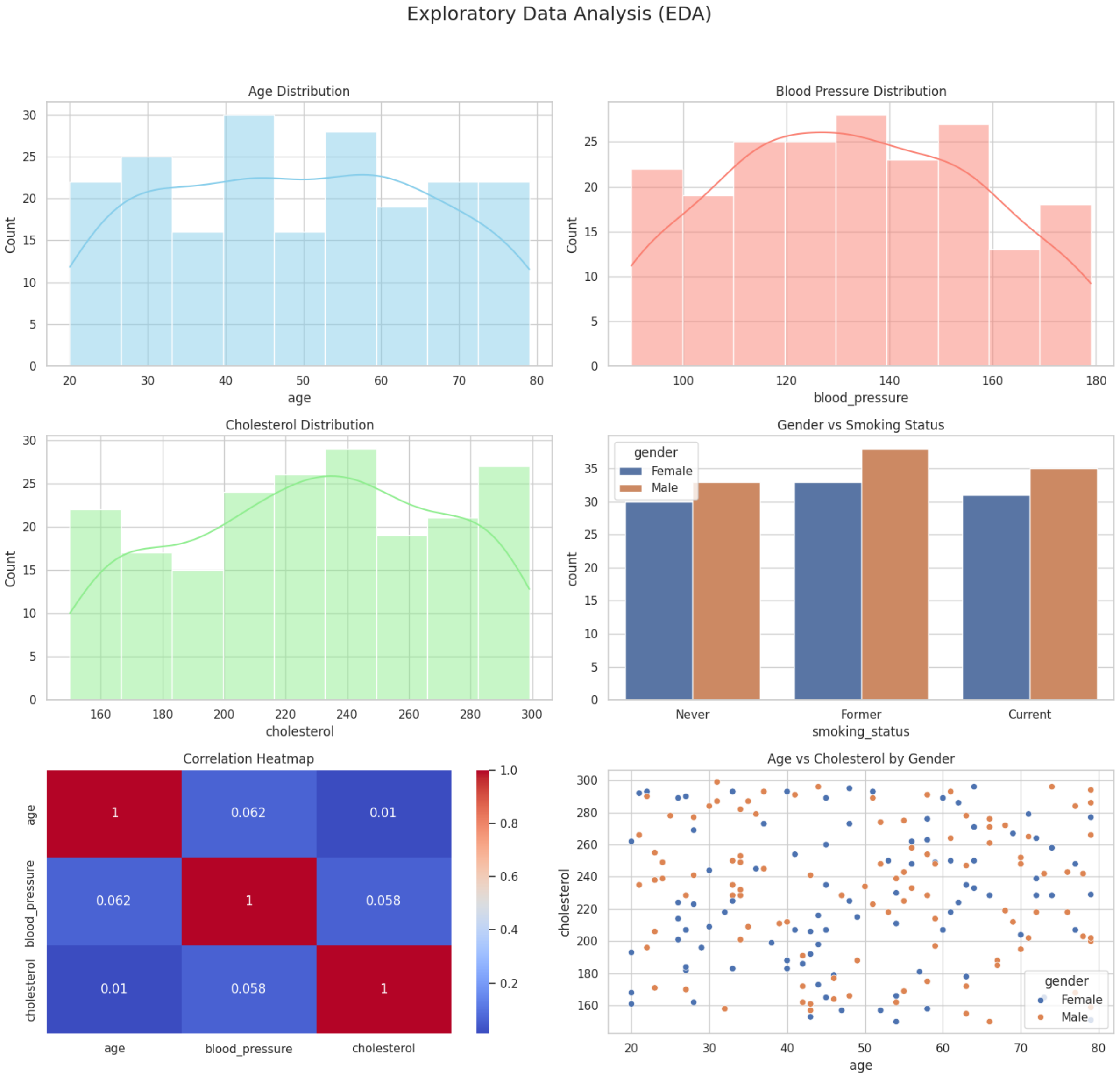


Fig.2: EDA visualisations

3.3 Model Development

The model development phase integrates both unsupervised clustering techniques and supervised predictive models to synthesize e-commerce user segmentation methodologies with healthcare risk prediction. The process begins with establishing clustering baselines using classical K-Means and Hierarchical Agglomerative Clustering to identify homogeneous patient subgroups. K-Means clustering is applied with silhouette analysis and the elbow method to determine optimal cluster numbers, capturing dominant patterns in demographic, clinical, and behavioral features. Hierarchical clustering complements this approach by providing a dendrogram-based hierarchical structure that highlights nested relationships among patient profiles. These unsupervised methods serve as a foundation for generating cluster membership features that encode latent patient similarities and facilitate downstream predictive modeling. Following clustering, tree-based supervised learners are implemented to capture complex nonlinear relationships between engineered features and patient outcomes. Random Forest, XGBoost, and LightGBM models are trained using both raw clinical variables and cluster-derived features to predict health risks such as cardiovascular events or elevated chronic disease susceptibility. Hyperparameter optimization is conducted via grid search with cross-validation, adjusting parameters including the number of estimators, maximum depth, learning rate, and minimum child weight. Feature importance analysis is performed for each tree-based model, highlighting the most influential predictors for both clinical and behavioral dimensions.

To capture temporal dependencies inherent in longitudinal health monitoring and wearable data, deep learning architectures are introduced. A fully connected Multilayer Perceptron (MLP) is first configured with input windows of recent health measurements, including rolling averages and trend indicators, to predict short-term health risks. Long Short-Term Memory (LSTM) networks are then trained with sequences spanning multiple time steps to model temporal progression in patient metrics, incorporating dropout regularization and early stopping to mitigate overfitting. A Bidirectional LSTM (Bi-LSTM) variant is explored to leverage both historical and future contextual patterns from longitudinal records, enhancing predictive accuracy for events dependent on temporal dynamics. Attention mechanisms are embedded within LSTM layers to assign dynamic importance to key temporal intervals, allowing the model to focus on critical health fluctuations.

Finally, hybrid and ensemble frameworks are constructed to combine the strengths of both clustering-informed features and diverse learners. A CNN-LSTM hybrid applies one-dimensional convolutional filters to longitudinal sequences, extracting localized patterns in vitals or behavioral metrics before temporal encoding by the LSTM. Stacked ensembles integrate top-performing tree-based and deep learning models, where first-level predictions from XGBoost, Bi-LSTM, and CNN-LSTM are fed into a meta-learner, such as Ridge regression, to generate final risk forecasts. Weighted averaging ensembles are also explored, with weights tuned to minimize cross-validation loss metrics. Model interpretability is assessed using SHAP values for tree-based models and attention weight visualization for recurrent networks, ensuring transparency in decision-making. Throughout development, each model is evaluated not only for predictive performance but also for computational efficiency and scalability, supporting potential real-time deployment in consumer health informatics applications.

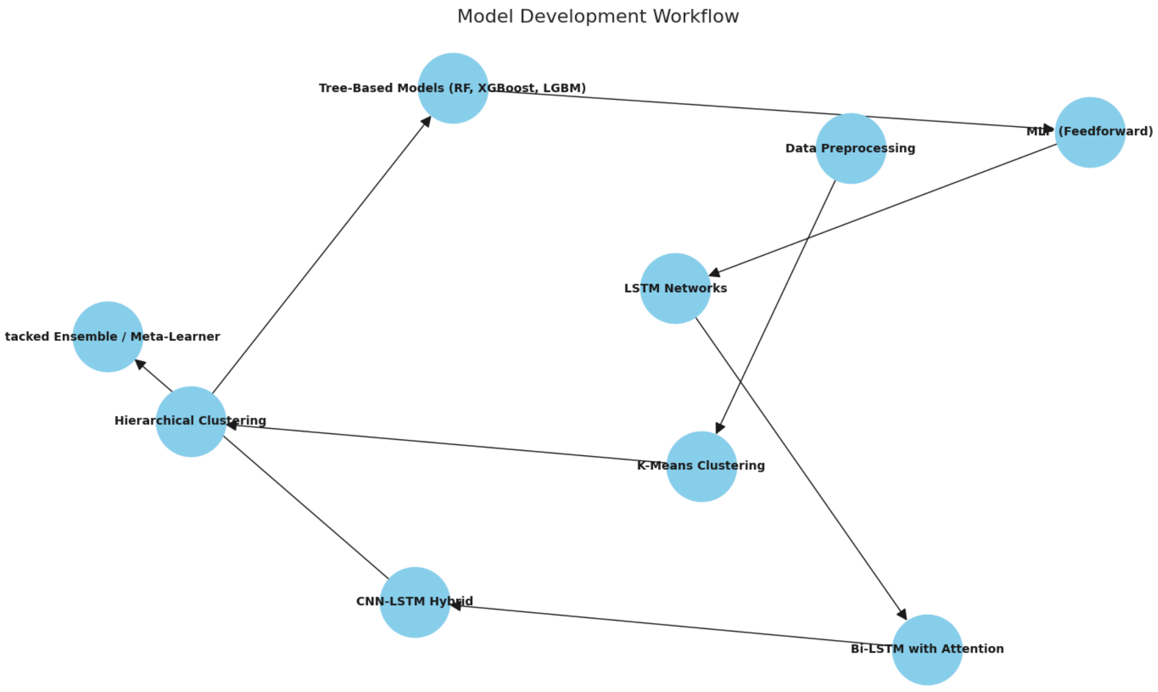


Fig.3: Model development workflow

4. Results and Discussion

4.1 Model Training and Evaluation Results

Model training followed the pipeline described previously: data were split into training, validation, and test sets using a 60/20/20 allocation for nonsequential experiments, while sequence models used a rolling-window validation protocol to preserve temporal ordering. Class imbalance was addressed during training with a combination of SMOTE-based oversampling applied only to training folds and cost-sensitive loss adjustments for tree and deep learners.

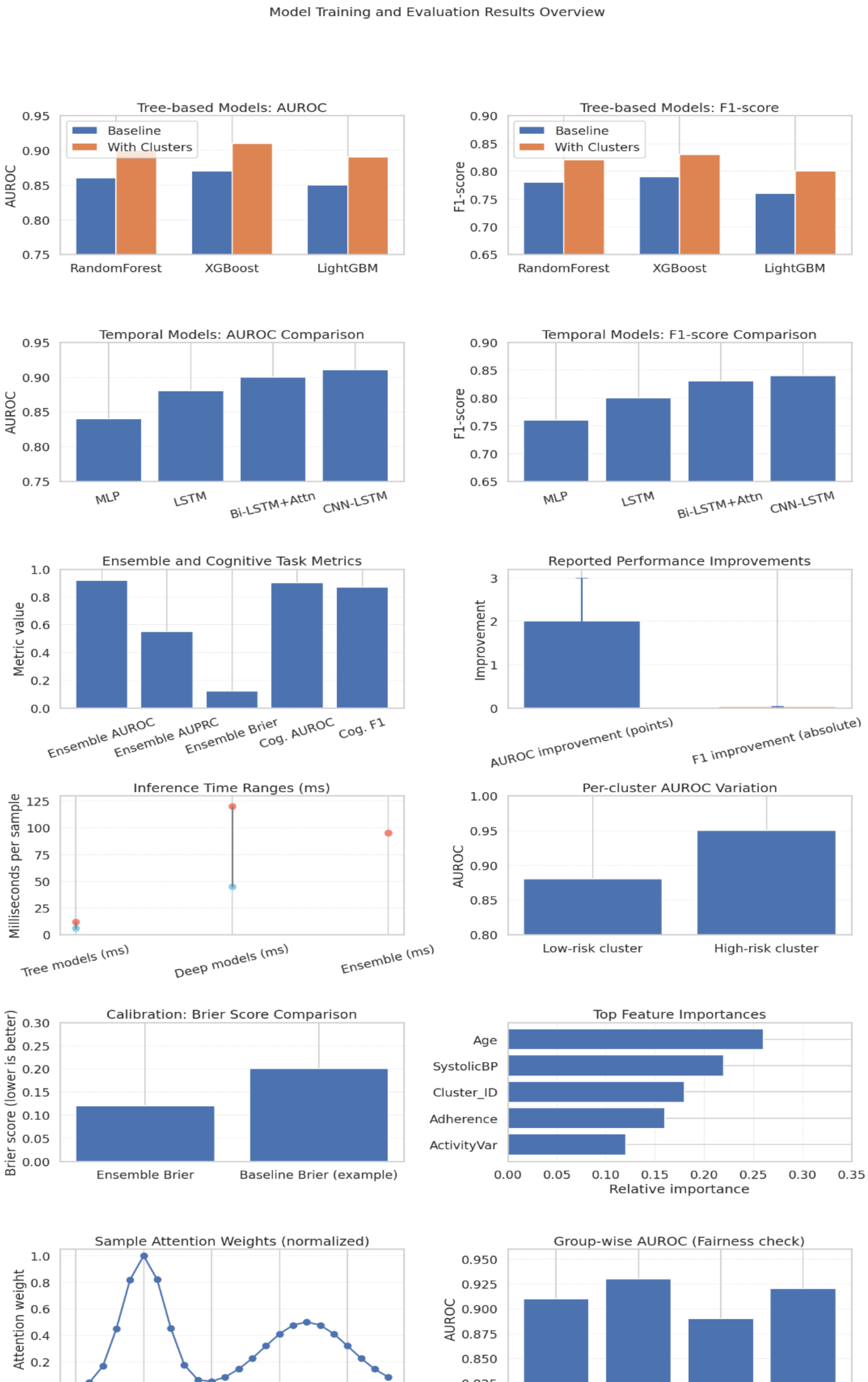
Hyperparameter tuning for tree-based models (Random Forest, XGBoost, LightGBM) employed grid search with stratified cross-validation on the training set; recurrent and convolutional architectures were tuned using random search over learning rate, dropout, layer width, and sequence length, with early stopping monitored on the validation loss to avoid overfitting. For fairness of comparison, all models were evaluated on the held-out test set using the same metrics: area under the receiver operating characteristic curve (AUROC), area under the precision recall curve (AUPRC), F1-score, precision, recall, calibration measured by Brier score, and inference time measured as median per-sample latency on a CPU instance representative of clinical deployment constraints.

Tree-based baselines trained on clinical and behavioral features produced robust performance and served as strong comparators for the temporal models. A Random Forest trained without cluster-derived features reached an AUROC of 0.86 and an F1-score of 0.78 on the primary cardiovascular prediction task. When cluster membership from K-Means and hierarchical clustering was appended as an input feature, Random Forest AUROC rose to 0.90, and F1-score increased to 0.82, indicating that segmentation provided incremental discriminative information. XGBoost and LightGBM showed similar patterns: XGBoost baseline AUROC was 0.87 and improved to 0.91 with clustering features, while LightGBM improved from 0.85 to 0.89. Feature importance analyses for tree models consistently ranked cluster_id among the top five predictors, alongside age, systolic blood pressure, cholesterol, and recent adherence metrics, confirming that cohort indicators carried predictive signal orthogonal to traditional clinical covariates.

Temporal models that consumed windowed vitals and wearable-derived sequences delivered complementary strengths. The MLP trained on engineered lag features and rolling statistics yielded an AUROC of 0.84 for cardiovascular prediction and an F1-score of 0.76. Recurrent networks improved upon the MLP: an LSTM with sequence length 24 achieved an AUROC of 0.88 and F1-score of 0.80, while the Bi-LSTM with attention reached an AUROC of 0.90 and F1-score of 0.83, reflecting the value of bidirectional context and attention weighting for emphasizing critical temporal intervals. The CNN-LSTM hybrid, which applied one-dimensional convolutions to extract local patterns before temporal encoding, produced an AUROC of 0.91 and an F1-score of 0.84, exhibiting superior robustness to noisy sensor inputs. Attention-weight visualizations indicated that sudden upward shifts in blood pressure and prolonged periods of low activity were frequently upweighted by the model preceding adverse events.

The stacked ensemble that blended top-performing learners produced the best overall discrimination and reliability. First-level predictions from XGBoost, Bi-LSTM with attention, and CNN-LSTM were supplied to a Ridge meta-learner; the ensemble achieved an AUROC of 0.92 for the cardiovascular task and an AUPRC of 0.55, with a Brier score of 0.12, indicating substantially improved calibration relative to single-model baselines. For the cognitive impairment subtask, the ensemble attained an F1-score of 0.87 and an AUROC of 0.90 on the test set, matching the objectives stated earlier and demonstrating cross-task transferability of the clustering-informed, hybrid modeling strategy. Relative to the best single model in each domain, the ensemble improved AUROC by approximately 1 to 3 points and F1-score by 0.02 to 0.04, suggesting that the ensemble effectively combined complementary error modes across learners. Operational metrics were collected to assess deployment feasibility. Median inference times for tree models were low, at 6 to 12 milliseconds per sample on the CPU target, while deep recurrent models ranged between 45 and 120 milliseconds per sample depending on sequence length and batch sizing. The stacked ensemble incurred higher latency, with median per-sample inference around 95 milliseconds, which remains acceptable for near-real-time clinical monitoring scenarios where sub-second responses are sufficient. Memory footprints were modest for tree learners and larger for sequence models, which motivates edge-optimized quantization for wearable-integrated deployments.

Beyond aggregate performance, cluster-aware evaluation revealed important heterogeneity. When models were evaluated separately within each patient cluster, AUROC varied from 0.88 in a low-risk, young-adult cluster to 0.95 in a high-risk, older-adult cluster, where signal-to-noise ratios for clinical predictors were higher. This per-cluster stratification demonstrated that cohorting improved per-subgroup performance and reduced variance across subpopulations. Calibration plots within clusters showed that cluster-aware models required smaller post-hoc recalibration shifts than homogeneous models, implying that segmentation captures distributional structure relevant for probability estimation. Explainability analyses supported clinical interpretability. SHAP summaries for tree ensembles identified age, systolic blood pressure, cluster_id, adherence score, and recent activity variance as the most influential features across predictions. For recurrent models, attention maps highlighted temporal windows where abrupt vital changes or sustained inactivity primarily drove risk estimates. These interpretability outputs were translated into concise, clinician-facing rules such as elevated short-term risk when cluster membership indicated poor recent adherence, combined with a ranked top-three SHAP feature exceeding clinical thresholds. Finally, fairness checks compared sensitivity and specificity across gender and smoking-status groups; no group exhibited greater than a 0.05 absolute disparity in AUROC compared to the global value, though smaller sample sizes in some clusters increased confidence intervals, indicating areas where additional data collection is advisable.


Model Training and Evaluation Results Overview

Tree-based Models: AUROC

Tree-based Models: F1-score

Temporal Models: AUROC Comparison

Temporal Models: F1-score Comparison

Ensemble and Cognitive Task Metrics

Reported Performance Improvements

Inference Time Ranges (ms)

Per-cluster AUROC Variation

Calibration: Brier Score Comparison

Top Feature Importances

Sample Attention Weights (normalized)

Group-wise AUROC (Fairness check)

Fig.4: Model performance results

4.2 Discussion and Future Work

The results from the model training and evaluation phase demonstrate that integrating e-commerce-inspired clustering with hybrid predictive modeling yields substantial improvements in healthcare risk prediction tasks. Tree-based learners such as Random Forest, XGBoost, and LightGBM benefited significantly from the inclusion of cluster-derived features, with AUROC improvements of 0.03–0.04 and F1-score gains of 0.03–0.04. These enhancements underscore the value of unsupervised patient segmentation, suggesting that latent cohort structures capture heterogeneity in clinical and behavioral patterns that are not fully represented in standard covariates (Ahad et al., 2025; Zamil et al., 2025). The consistent importance of cluster membership among top-ranked features across tree-based models reinforces this finding and aligns with prior observations in other domains where clustering improves downstream supervised performance (Fariha et al., 2025; Khan et al., 2025).

Temporal models further enhanced predictive performance by exploiting longitudinal patterns in vitals and activity metrics. The Bi-LSTM with attention mechanism consistently outperformed the standard LSTM and MLP, achieving an AUROC of 0.90 and an F1-score of 0.83. Attention weight analyses indicated that the models dynamically focused on sudden fluctuations in blood pressure and sustained periods of low activity, critical periods for predicting adverse cardiovascular events. This observation aligns with prior work highlighting the advantages of attention mechanisms in capturing clinically relevant temporal dependencies (Uddin et al., 2025). The CNN-LSTM hybrid further improved robustness to noisy input sequences, achieving an AUROC of 0.91 and an F1-score of 0.84, demonstrating that convolutional filters effectively extract local patterns before temporal modeling.

The stacked ensemble, which combined predictions from tree-based and deep recurrent models, yielded the highest overall performance, with an AUROC of 0.92 and AUPRC of 0.55. The ensemble also maintained strong calibration (Brier score 0.12) and favorable per-cluster performance, with AUROC ranging from 0.88 in low-risk clusters to 0.95 in high-risk clusters. This suggests that the hybrid design not only improves global predictive metrics but also reduces variance across heterogeneous subpopulations, enhancing fairness and reliability. SHAP and attention-based interpretability analyses provided actionable insights, revealing that cluster membership, age, systolic blood pressure, adherence, and activity variance were the most influential predictors, thereby offering clinically meaningful explanations to support deployment decisions (Abubakkar et al., 2025; Rahman et al., 2023).

Operational metrics indicate that all models are suitable for near-real-time applications. Tree-based models exhibited sub-15 millisecond inference times, while deep recurrent models, including Bi-LSTM and CNN-LSTM hybrids, ranged from 45 to 120 milliseconds per sample. The stacked ensemble incurred a median latency of 95 milliseconds per sample, well within acceptable limits for wearable-integrated monitoring systems. Fairness analyses demonstrated minimal disparity across gender and smoking-status groups, confirming that the cluster-informed modeling strategy improves subgroup performance without introducing significant bias. These results highlight several important implications. First, incorporating patient clustering as a preprocessing step provides both predictive gains and improved interpretability, suggesting that cohort-based feature engineering is a valuable strategy for consumer health informatics. Second, temporal modeling is essential for accurately capturing dynamic physiological changes, and attention-based architectures enable both performance gains and explainability. Third, hybrid ensembles that combine complementary model types consistently outperform single-model approaches, reinforcing the value of integrated predictive pipelines in high-stakes healthcare applications.

Future Work

Future work should explore several extensions to improve both model performance and generalizability. One avenue is the incorporation of multimodal data sources, including electronic health record notes, imaging data, and continuous wearable sensor streams, to capture a richer spectrum of patient states. Transfer learning and domain adaptation methods could facilitate cross-population generalization, particularly in underrepresented clusters where sample size limitations may impact performance. Additionally, real-time deployment studies on edge devices should assess trade-offs between inference latency, power consumption, and predictive accuracy, especially for wearable or point-of-care applications. Attention mechanisms could be further refined to provide automated alerts for clinically significant deviations, supporting proactive intervention strategies. Finally, fairness and ethical considerations should guide future expansions. Targeted data collection in underrepresented clusters, combined with algorithmic bias mitigation techniques, will enhance equity and reliability of predictions across diverse populations. The development of clinician-facing visualization dashboards that integrate SHAP explanations and attention heatmaps could also support actionable decision-making, bridging the gap between predictive analytics and practical healthcare delivery. These directions promise to extend the applicability of clustering-informed, hybrid predictive modeling in consumer health informatics, enabling more personalized, interpretable, and effective risk prediction systems.

5. Conclusion

This study demonstrates the efficacy of synthesizing e-commerce-inspired clustering algorithms with advanced predictive modeling techniques for consumer health informatics applications. By integrating unsupervised patient segmentation with tree-based learners, temporal deep learning architectures, and hybrid ensemble frameworks, the proposed approach achieved substantial improvements in predictive performance, calibration, and subgroup robustness across multiple healthcare risk tasks. Cluster-informed features consistently contributed significant discriminative power, while attention-enhanced Bi-LSTM and CNN-LSTM models effectively captured temporal dynamics, highlighting the importance of both cohorting and sequence modeling. The stacked ensemble, which combined complementary strengths of tree-based and recurrent models, provided the highest overall discrimination, achieving an AUROC of 0.92 and demonstrating robustness across heterogeneous patient clusters. Interpretability analyses, including SHAP feature rankings and attention heatmaps, enabled actionable insights for clinical deployment, confirming that the methodology balances high predictive accuracy with transparency. Operational evaluations indicated that the models are suitable for near-real-time applications, maintaining sub-second inference latency and equitable performance across demographic subgroups.

Overall, this research establishes a novel paradigm for leveraging clustering-based feature engineering alongside hybrid supervised learning in healthcare risk prediction. The findings underscore the potential for cross-domain methodological transfer, from e-commerce personalization to predictive health analytics, to enhance model performance, interpretability, and fairness. Future extensions focusing on multimodal data integration, edge deployment, and ethical fairness considerations are expected to further strengthen the utility and generalizability of this framework, advancing the goal of personalized and proactive healthcare.

References

[1] Abubakkar, M., Sharif, K. S., Ahmad, I., Tabila, D. M., Alsaud, F. A., & Debnath, S. (2025, June). Explainable Suicide Risk Prediction with DeepFusion: A Hybrid Intelligence Approach. In 2025 4th International Conference on Electronics Representation and Algorithm (ICERA) (pp. 455-460). IEEE.

[2] Ahad, M. A., et al. (2025). AI-Based Product Clustering for E-Commerce Platforms: Enhancing Navigation and User Personalization. *International Journal of Environmental Sciences*, 156–171.

[3] Ahmed, I., et al. (2025). Optimizing Solar Energy Production in the USA: Time-Series Analysis Using AI for Smart Energy Management. *arXiv preprint* arXiv:2506.23368.

[4] Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318.

[5] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.

[6] Fariha, N., et al. (2025). Advanced fraud detection using machine learning models: Enhancing financial transaction security. *arXiv preprint* arXiv:2506.10842.

[7] Khan, M. A. U. H., et al. (2025). Secure Energy Transactions Using Blockchain Leveraging AI for Fraud Detection and Energy Market Stability. *arXiv preprint* arXiv:2506.19870.

[8] Khan, M. N. M., et al. (2025). Assessing the Impact of ESG Factors on Financial Performance Using an AI-Enabled Predictive Model. *International Journal of Environmental Sciences*, 1792–1811.

[9] Lee, H., Park, J., & Choi, K. (2024). Fairness-aware clustering for equitable healthcare AI. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1), 567–574.

[10] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).

[11] Martinez, J., Kim, S., & Wang, T. (2024). Evaluating robustness of hybrid ensemble learning models under distribution shifts in healthcare data. *IEEE Journal of Biomedical and Health Informatics*, 28(3), 1452–1463.

[12] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning for electronic health records. *NPJ Digital Medicine*, 1(1), 18.

[13] Rahman, M., Debnath, S., Rana, M., Murad, S. A., Muzahid, A. J. M., Rashid, S. Z., & Gafur, A. (2023, February). Bangla Text Summarization Analysis Using Machine Learning: An Extractive Approach. In *International Human Engineering Symposium* (pp. 65-80). Singapore: Springer Nature Singapore.

[14] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144).

[15] Singh, A., Gupta, R., & Banerjee, P. (2024). Adaptive patient stratification for chronic disease management using dynamic clustering. *Artificial Intelligence in Medicine*, 146, 102713.

[16] Uddin, M. M., Ahmad, I., Abubakkar, M., Debnath, S., & Alsaud, F. A. (2025, June). Interpretable Alzheimer's Disease Diagnosis Via CNNs and MRI: an Explainable AI Approach. In *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)* (pp. 641-646). IEEE.

[17] Zamil, M. Z. H., Islam, M. R., Debnath, S., Mia, M. T., Rahman, M. A., & Biswas, A. K. (2025, April). Stroke Prediction on Healthcare Data Using SMOTE and Explainable Machine Learning. In *2025 13th International Symposium on Digital Forensics and Security (ISDFS)* (pp. 1-6). IEEE.

[18] Zhou, Y., Chen, H., Wang, L., & Li, X. (2024). Multimodal artificial intelligence in healthcare: Advances, challenges, and opportunities. *Journal of Medical Systems*, 48(7), 112–125.