Journal of Information Science Research and Practice

Research Articles

การศึกษาการทำนายความเสี่ยงของการเกิดโรคร่วมของโรคเรื้อรังของผู้สูงอายุในชนบทของจีนโดยใช้การเรียนรู้เชิงลึกแบบหลายป้ายกำกับ

Study on Risk Prediction of Comorbidity of Major Chronic Diseases in Rural Elderly in China Based on Multi-label Deep Learning

เฟิน ฟู [1] กิตติยา สุทธิประภา [1,*] และ กันยารัตน์ เควียเซ่น [1]

Fen Fu [1] Kittiya Suthiprapa [1,*] and Kanyarat Kwiecien [1]

[1] สาขาวิชาสารสนเทศศาสตร์ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น ประเทศไทย; Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand

* Corresponding author email: kittsu@kku.ac.th

บทคัดย่อ

วัตถุประสงค์: เพื่อพัฒนาและประเมินกรอบการเรียนรู้เชิงลึกที่มีหลายป้ายกำกับสำหรับการทำนายโรคเรื้อรังที่สำคัญห้าโรคไปพร้อมๆ กัน

วิธีการศึกษา: การใช้ข้อมูล CHARLS ที่เสริมด้วยบันทึกการตรวจสุขภาพในชนบท (n=38,569, อายุ ≥60) เครือข่ายประสาทเชิงลึกแบบหลายป้ายกำกับที่เชื่อมต่ออย่างสมบูรณ์ได้รับการฝึกอบรมด้วยการเพิ่มประสิทธิภาพเกณฑ์เอนโทรปีข้ามไบนารี่ถ่วงน้ำหนักและการปรับเกณฑ์เฉพาะโรค มีการใช้การตรวจสอบความถูกต้องข้ามห้าเท่าและชุดการทดสอบอิสระ (15%) สำหรับการประเมิน

ข้อค้นพบ: โมเดลดังกล่าวได้รับ AUC-ROC ระดับมาโครที่ 0.8587 และคะแนนมาโคร F1 ที่ 0.6676 ในชุดการทดสอบอิสระ ซึ่งมีประสิทธิภาพเหนือกว่า Logistic Regression และ XGBoost ในขณะที่บรรลุประสิทธิภาพการแข่งขันเมื่อเปรียบเทียบกับฟอเรสต์สุ่ม การวิเคราะห์ SHAP ระบุความดันโลหิตซิสโตลิก BMI และระดับน้ำตาลในเลือดขณะอดอาหารเป็นตัวทำนายอันดับต้นๆ

การประยุกต์ใช้จากการศึกษานี้: กรอบการทำงานที่นำเสนอแสดงให้เห็นถึงความเป็นไปได้ในการเปลี่ยนแปลงข้อมูลการสำรวจระดับประชากรเป็นเครื่องมือที่มีความเสี่ยงต่อโรคหลายโรคที่สามารถนำไปปฏิบัติได้สำหรับบริการปฐมภูมิในชนบทที่มีทรัพยากรจำกัด

คำสำคัญ: การเรียนรู้เชิงลึก การเรียนรู้แบบหลายป้ายกำกับ การทำนายความเสี่ยงของโรคเรื้อรัง ผู้สูงอายุในชนบท CHARLS

Received: February 10, 2026 Revised: March 28, 2026 Accepted: March 30 , 2026 https://doi.org/10.14456/jiskku.2026.4

Abstract

Purpose: To develop and evaluate a multi-label deep learning framework for simultaneously predicting five major chronic diseases.

Methodology: Using CHARLS data supplemented with rural health examination records (n=38,569, aged ≥60), a fully connected multi-label deep neural network was trained using weighted binary cross-entropy and disease-specific threshold optimization. Five-fold cross-validation and an independent test set (15%) were used for evaluation.

Findings: The model achieved a macro AUC-ROC of 0.8587 and a macro F1-score of 0.6676 on the independent test set, outperforming logistic regression and XGBoost while achieving competitive performance compared to random forest. SHAP analysis identified systolic blood pressure, BMI, and fasting glucose as the top predictors.

Applications of this study: The proposed framework demonstrates the feasibility for transforming population-level survey data into actionable multimorbidity risk tools for resource-constrained rural primary care.

Keywords: Deep learning, Multi-label learning, Chronic disease risk prediction, Rural older adults, CHARLS

1. Introduction

Population aging and the increasing burden of chronic diseases pose substantial challenges to primary healthcare systems worldwide (WHO, 2025; UNDESA, 2022). According to the United Nations World Population Prospects 2022, the global population aged 60 years and above is projected to reach 2.1 billion by 2050, accounting for roughly 22% of the world's population (UNDESA, 2022). China, which has the largest number of older adults worldwide, is undergoing rapid and large-scale aging. By the end of 2024, China had 310 million people aged ≥60 years (22.0%) and 220 million aged ≥65 years (15.6%) (National Bureau of Statistics of China, 2025; Zhang, 2025). The aging trend is particularly pronounced in rural areas, where the share of rural residents aged ≥60 years reached 23.81%, about 7.99 percentage points higher than in urban areas (National Bureau of Statistics of China, 2021). This "rural–urban inversion," combined with the phenomenon of "aging before affluence," places considerable pressure on rural healthcare systems, chronic disease management, and long-term care

(Ministry of Civil Affairs, National Development and Reform Commission, Ministry of Finance, 2025).

Chronic noncommunicable diseases have become the primary health threat to rural older adults. The World Health Organization estimates that noncommunicable diseases account for approximately 71% of global deaths (WHO, 2025). In China, chronic conditions are highly prevalent among rural older adults, and multimorbidity is common in later life (Zhang et al.,2022). A rural-based cross-sectional study reported substantial prevalence of chronic conditions and frequent multimorbidity among older adults, underscoring the limitations of traditional single-disease prediction models and the need for tools capable of simultaneously assessing multiple disease risks (Wang et al, 2015).

Rural healthcare systems face significant structural barriers in meeting the needs of chronic disease management. Previous studies have reported spatiotemporal disparities in China's healthcare resource allocation, with lower resource levels in less-developed rural regions (Jianpeng, 2022). Early identification and standardized management of chronic diseases remain persistent challenges in rural primary care. Gaps in screening, follow-up, and adherence to standardized management pathways are more prominent in rural areas than in urban settings, contributing to delayed recognition and missed intervention windows (Ministry of Civil Affairs, National Development and Reform Commission, Ministry of Finance, 2025). Risk prediction tools are still limited in routine grassroots practice, and decision-making often relies on clinical experience or simple rule-based screening rather than systematic data-driven evaluation. Traditional risk assessment approaches may fail to exploit multidimensional health data, and substantial underdiagnosis persists in routine care, as evidenced by recent observational studies documenting delayed recognition of chronic conditions in primary care settings (Nie et al, 2025). Zhang et al. (Chen et al, 2024) highlighted significant deficits in the capacity of rural primary care physicians to provide geriatric services, including shortages of trained personnel, limited professional skills, outdated equipment, and low levels of digital health adoption.

From a methodological perspective, deep learning has rapidly advanced from theoretical exploration to wide-ranging applications in healthcare. The landmark review by LeCun et al. (LeCun, Bengio &Hinton, 2024) published in Nature established the conceptual and methodological foundation for deep learning, highlighting its superior capability in modeling complex, high-dimensional data. In recent years, deep learning has achieved notable breakthroughs in medical imaging, disease risk prediction, and drug discovery. Esteva et al. (Esteva et al, 2017) demonstrated dermatologist-level performance in skin cancer classification using deep convolutional neural networks (AUC = 0.96), while Gulshan et al (2016) developed a deep learning system for diabetic retinopathy screening that achieved both sensitivity and specificity exceeding 90%.

Deep learning has also demonstrated strong capabilities in predicting chronic disease risks using structured clinical data. Rajkomar et al. 2018) showed that deep learning models significantly outperformed traditional approaches across multiple clinical outcomes based on electronic health records from 216,221 patients. Substantial progress has been made specifically in cardiovascular disease prediction. Weng et al. (2014) used neural networks to predict cardiovascular disease risk, achieving higher accuracy than guideline-recommended risk scores from the American College of Cardiology/American Heart Association (ACC/AHA), while Alaa et al. (Alaa et al, 2019) developed an automated machine learning framework for cardiovascular disease risk prediction in 423,604 UK Biobank participants. Choi et al. (2014) introduced the Doctor AI model utilizing recurrent neural networks to predict future diagnoses with strong performance (AUC = 0.83 for heart failure), and Miotto et al. (2016) proposed the Deep Patient model achieving superior predictive accuracy across multiple chronic diseases (Dai et al, 2024).

More recently, transformer-based architectures have demonstrated strong capabilities in predicting disease risk and progression from longitudinal electronic health records. Zisser and Aran (2024). developed STRAFE, a transformer-based survival analysis architecture for EHR data, which outperformed conventional time-to-event prediction algorithms in predicting chronic kidney disease deterioration across a real-world dataset of over 130,000 individuals. Building on pre-trained EHR foundation models, He et al. (2025). refined Med-BERT by reformulating disease prediction as a next-visit token prediction task, achieving improved performance in pancreatic cancer prediction under both few-shot and fully supervised settings. These advances underscore the growing potential of transformer-based deep learning for clinical disease prediction using longitudinal health data. However, their computational complexity and reliance on large-scale longitudinal EHRs may limit deployment in resource-constrained rural primary care settings—a challenge directly addressed by the lightweight fully connected architecture proposed in the present study.

Multi-label learning is a machine learning paradigm in which each instance can be associated with multiple labels simultaneously (Tsoumakas & Katakis, 2007). In the medical field, this approach is particularly suitable for modeling multimorbidity, where patients often present with several coexisting chronic conditions. Multi-label models enable simultaneous prediction of multiple diseases while capturing interdependencies among them (Tsoumakas & Katakis & Vlahavas, 2009). Despite these benefits, multi-label deep learning research focused on chronic diseases in rural older adults remains limited. Existing studies typically address single-disease prediction or simple binary classification tasks, failing to capture the multimorbidity patterns characteristic of elderly populations.

Recent studies have demonstrated the growing utility of multi-label learning frameworks in clinical disease prediction. El-Hasnony et al. (2022). applied multi-label active learning with five selection strategies to predict heart disease risk, demonstrating that multi-label approaches achieve superior generalization compared to conventional single-label classifiers and effectively reduce labelling costs in healthcare settings. More broadly, advanced deep learning architectures have shown strong potential for simultaneous multi-disease risk assessment using routinely available data. Dong et al. (2025) developed Omicsformer, a deep learning model integrating multi-omics and routine blood test data, achieving accurate risk prediction for nine chronic diseases including cardiovascular and metabolic conditions, validated across large-scale electronic health records. These advances collectively reinforce the suitability of multi-label deep learning frameworks for the multimorbidity prediction context addressed in the present study.

Based on existing literature, two key gaps remain insufficiently addressed. First, the application of deep learning methods for chronic disease risk prediction among rural older adults is still limited, especially when relying on routinely collected primary care data rather than high-cost hospital-based EHRs. Second, despite the high prevalence of multimorbidity in aging populations, few studies have adopted multi-label learning frameworks capable of simultaneously modeling multiple chronic diseases and their interdependencies in a unified predictive model.

Unlike prior single-disease models (e.g., (Wang et al, 2015; Alaa et al, 2019; Liang et al, 2025)) that independently predict hypertension, diabetes, or cardiovascular disease in isolation, our framework jointly learns shared and disease-specific feature representations, captures inter-disease correlations within a unified architecture, and produces simultaneous multi-disease risk scores. This design is fundamentally different from training five separate binary classifiers and addresses the real-world clinical scenario where older adults present with overlapping risk profiles.

To address these gaps, this study aims to: (1) develop a multi-label deep learning framework to simultaneously predict the risks of five major chronic diseases among rural older adults; (2) evaluate the model's predictive performance and robustness using rigorous cross-validation and an independent test set; (3) enhance model transparency through SHAP-based interpretability to support clinical trust and practical adoption; and (4) demonstrate how population-based survey data can be transformed into actionable risk prediction tools suitable for resource-constrained rural primary care settings.

The main contributions of this study are summarized as follows:

(1) Dataset Construction: We constructed a comprehensive multi-label prediction dataset by integrating CHARLS rural samples with external rural health examination records, covering five major chronic diseases through systematic harmonization (n=38,569).

(2) Model Development: We designed a three-layer fully connected multi-label deep neural network with weighted binary cross-entropy loss and disease-specific threshold optimization, validated against logistic regression, random forest, and XGBoost baselines.

(3) Interpretability: We integrated SHAP-based feature attribution analysis to identify clinically plausible predictor contributions, enhancing model transparency and clinical trust.

(4) Practical Pathway: We demonstrated a system-ready pathway for embedding multi-label risk prediction models into rural health information systems, providing methodological guidance for future deployment in resource-constrained settings.

2. Purpose

This study aimed to develop a deep learning–based system for managing health records and predicting disease risks among rural elderly populations in China.

3. Methodology

3.1 Data Sources and Study Population

This retrospective cohort study utilized data from the China Health and Retirement Longitudinal Study (CHARLS), a nationally representative survey covering individuals aged 45 years and older (Zhao et al.,2016), supplemented with de-identified rural physical examination records obtained through academic collaboration. CHARLS provides high-quality, multidimensional microdata suitable for predictive modeling research involving older adults.

The survey collects comprehensive demographic, socioeconomic, health status, and healthcare utilization information through standardized questionnaires and physical examinations.

An additional dataset was obtained from an elderly physical examination program conducted in rural regions under China's Basic Public Health Service Program (Ministry of Civil Affairs, National Development and Reform Commission, Ministry of Finance, 2025). These examinations are carried out annually by community and township health centers. The external dataset closely matched CHARLS in questionnaire structure, health indicators, chronic disease definitions, and sampling population characteristics, facilitating data integration and harmonization.

CHARLS data were collected with ethical approval from the Peking University Biomedical Ethics Committee (IRB00001052-11015), and all respondents provided informed consent (Zhao et al.,2016). The external dataset was fully de-identified before use, with all personal identifiers removed to protect participant privacy.

The target population consisted of rural older adults aged ≥60 years. Exclusion criteria included: (1) missing age, sex, or outcome variables; (2) predictor variables with >30% missingness; and (3) logical errors such as negative values or inconsistent demographic information. After data cleaning, feature harmonization, and diagnostic criteria standardization, the CHARLS data were merged with the external rural examination dataset. The final analytical dataset comprised 38,569 eligible individuals with 21 variables: 16 predictors and 5 chronic disease outcome variables. The supplementary dataset was used to mitigate extreme class imbalance rather than to alter population-level prevalence, and all analyses were conducted under the assumption that the combined dataset represents a pragmatic screening population rather than a nationally weighted sample.

3.2 Outcome Variables and Disease Definitions

This study simultaneously predicted five common chronic diseases: hypertension, diabetes, heart disease, dyslipidemia, and stroke. These diseases were selected because they have high prevalence among rural elderly populations, are major contributors to mortality and disability in older adults (WHO, 2025) and share modifiable risk factors that can inform preventive interventions.

Disease labels were derived using a comprehensive approach combining self-reported physician diagnoses, medication history, and physical examination indicators. The definitions followed standard clinical guidelines and existing empirical research (Wang et al, 2015). All outcome variables were coded as binary labels: 1 indicating presence of a diagnosis or abnormal indicator, and 0 indicating absence of diagnosis with indicators within the normal range.

Preliminary analysis of the CHARLS rural sample revealed moderate to severe class imbalance for certain diseases, particularly diabetes (prevalence: 9.2%) and stroke (prevalence: 4.8%). Class imbalance can lead to biased model training, where the model favors the majority class and performs poorly on minority classes, reducing sensitivity for detecting at-risk individuals.

To address this issue, we supplemented the CHARLS dataset with external rural physical examination data that closely matched the CHARLS sampling framework, questionnaire design, and diagnostic criteria. After integration, disease prevalence increased moderately while maintaining realistic distribution patterns (see Figure 1).

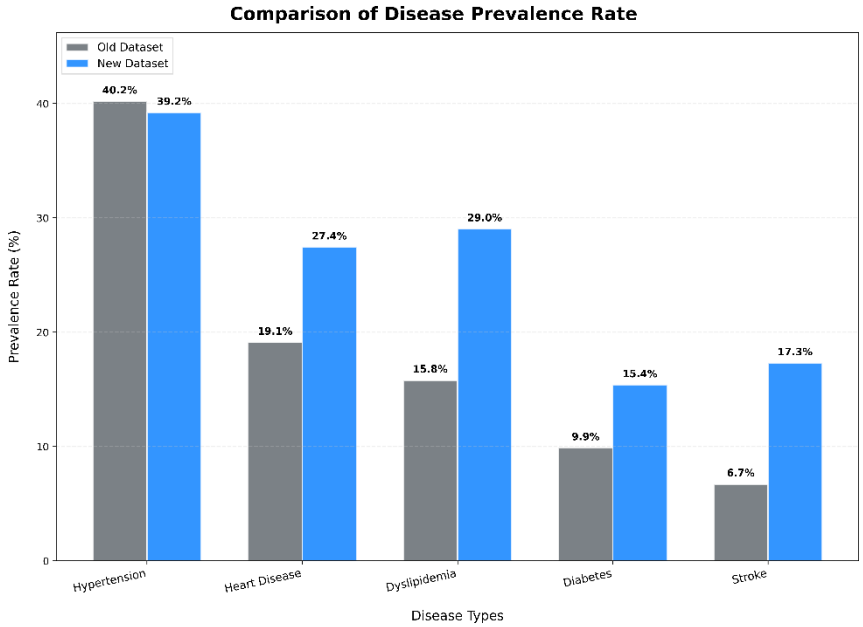


Figure 1 Comparison of disease prevalence before and after supplementation

3.3 Predictor Variables

Sixteen predictor variables were selected based on clinical relevance, data availability, and alignment with routine primary care practice. The predictors included demographic characteristics (age, sex, region of residence), anthropometric measurements (body mass index), vital signs (systolic and diastolic blood pressure, pulse rate), laboratory test results (fasting plasma glucose, HDL cholesterol, LDL cholesterol, total cholesterol), socioeconomic indicators (total household income, proportion of household chronic disease expenses), and self-reported health status (self-rated health, activities of daily living, cognitive function).

These indicators were selected because they: (1) are routinely collected during primary care visits and annual health examinations; (2) have established associations with chronic disease risk based on epidemiological evidence (Zhang et al., 2022; Wang et al., 2015); (3) are low-cost and widely accessible, making them feasible for resource-constrained rural settings; and (4) cover multiple risk domains, including biological, behavioral, and socioeconomic factors.

Importantly, the selected predictors do not require advanced diagnostic equipment or specialized laboratory tests, enhancing the model's practical applicability for rural primary care. This design philosophy prioritizes scalability and real-world deployment over maximizing predictive accuracy through inclusion of costly or unavailable biomarkers.

Table 1 Predictor variables and clinical interpretations

NO.	FEATURE	FULL NAME	CLINICAL MEANING
1	age	Age (years)	A primary non-modifiable risk factor; chronic disease risk increases significantly with age.
2	city	Region of residence	Reflects geographic, economic, and healthcare access disparities influencing disease distribution.
3	bmi	Body Mass Index (kg/m^2)	Indicator of nutritional status and obesity; closely linked to diabetes, hypertension, and dyslipidemia.
4	systo	Systolic Blood Pressure (mmHg)	Core diagnostic criterion for hypertension; elevated levels increase risk of heart disease and stroke.
5	diasto	Diastolic Blood Pressure (mmHg)	Reflects vascular elasticity and cardiovascular function; chronic elevation is a risk factor for heart failure and stroke.

NO.	FEATURE	FULL NAME	CLINICAL MEANING
6	pulse	Pulse Rate (beats/min)	Indicates cardiac output and autonomic nervous system regulation; abnormal values correlate with cardiovascular diseases.
7	bl_glu	Fasting Plasma Glucose (mmol/L)	Direct diagnostic measure for diabetes; reflects pancreatic beta-cell function and metabolic control.
8	bl_hdl	HDL Cholesterol (mmol/L)	"Good cholesterol"; low levels increase the risk of atherosclerosis.
9	bl_ldl	LDL Cholesterol (mmol/L)	"Bad cholesterol"; elevated levels are a major risk factor for coronary heart disease and stroke.
10	bl_cho	Total Cholesterol (mmol/L)	Reflects overall lipid metabolism status; a key indicator for dyslipidemia.
11	income_total	Total Household Income	Indicator of socioeconomic status; strongly associated with dietary patterns and healthcare-seeking behavior.
12	hhcperc	Proportion of Household Health Expenditure	Indicator of healthcare burden and financial stress; indirectly reflects health management capacity.
13	cesd10	CESD-10 Score (Depression)	Depression is highly comorbid with chronic conditions; mental health affects lifestyle and treatment adherence.
14	total_cognition	Global Cognitive Score	Indicator of brain health in the elderly; associated with sequelae of stroke and cardiovascular diseases.
15	totmet	Total MET (Physical Activity Level)	Reflects occupational/leisure-time physical activity; physical inactivity is a major risk factor for chronic diseases.
16	srh	Self-Rated Health	Captures an individual's holistic health perception; a strong predictor of mortality and morbidity risk.

3.4 Data Preprocessing

Data preprocessing is a critical step in the machine learning modeling process, directly impacting model performance and stability. This study employs a standardized preprocessing workflow comprising three main steps: missing value handling, feature scaling, and dataset partitioning.

For missing value handling, we first conducted a missing mechanism analysis. Through Little's MCAR (Missing Completely At Random) test (χ^2=142.3, p=0.18), this result supports the assumption that the data's missing patterns are consistent with completely random missingness, rather than conclusively proving MCAR, providing statistical justification for employing simple imputation methods. Features were systematically categorized: clinical indicators (age, BMI, blood pressure, blood glucose, blood lipids), cognitive function, and socioeconomic variables were classified as numerical features, while geographic information was treated as a categorical feature. For numerical variables, missing values were imputed using median imputation followed by standardization scaling. For categorical variables, mode imputation combined with one-hot encoding conversion was applied, effectively handling mixed data types while preserving distribution stability.

Feature scaling is critical for deep learning models, as differing dimensions and numerical ranges of features may destabilize gradient descent. For all numerical variables, we employ Z-score standardization, transforming each feature into a distribution with mean 0 and standard deviation 1. The normalization formula is: x_scaled = ($\boldsymbol{x}$ - $\boldsymbol{\mu}$) / $\boldsymbol{\sigma}$, where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ represent the mean and standard deviation of the feature in the training set, respectively. It is important to note that the test set must be normalized using $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ calculated from the training set to avoid introducing forward-looking information. For categorical variables (in this study, only residential area), we employed one-hot encoding to convert them into binary vector representations suitable for neural network processing.

The dataset was partitioned using stratified random sampling, dividing the consolidated 38,569 cases into a training set (32,783 cases) and an independent test set (5,786 cases) at an 85:15 ratio. Hypertension prevalence served as the stratification variable for hierarchical sampling, as it exhibited the highest prevalence and representativeness among the five diseases. This strategy ensured consistent disease distribution between training and test sets relative to the population, mitigating sampling bias inherent in random selection. The test set was strictly maintained as independent throughout the entire model development process, not participating in any model training or hyperparameter tuning. It was used only during the final model evaluation stage to provide an unbiased estimate of the model's generalization capability.

3.5 Model Development and Architecture

A multi-label fully connected deep neural network was constructed to simultaneously predict the risk of five chronic diseases (LeCun & Bengio, Hinton, 2024; Tsoumakas & Katakis, 2007; (Tsoumakas & Katakis & Vlahavas, 2009). The architecture consisted of an input layer, three hidden layers, and an output layer. The input layer accepted 16 predictor variables. The three hidden layers contained 256, 128, and 64 neurons respectively, with progressive dimensionality reduction to extract hierarchical feature representations. Rectified Linear Unit (ReLU) activation functions were used in hidden layers to introduce non-linearity and enable learning of complex patterns. Dropout layers with a rate of 0.5 were inserted after each hidden layer to prevent overfitting and improve generalization.

The output layer contained 5 neurons corresponding to the five chronic diseases. Sigmoid activation functions were used in the output layer to generate independent probability estimates for each disease, ranging from 0 to 1. This design enabled the model to output multiple binary predictions simultaneously, with each disease treated as an independent binary classification task while sharing the same feature representation learned by the hidden layers.

The model was implemented using TensorFlow 2.x and Keras. The network architecture was designed to balance model capacity and overfitting risk. Three hidden layers were selected based on preliminary experiments showing that deeper architectures did not yield substantial performance gains and increased training time. The neuron counts (256-128-64) provided sufficient capacity to capture complex non-linear relationships while maintaining computational efficiency.

Model Training and Optimization Details: The model was trained using the Adam optimizer with an initial learning rate of 0.0005 and a batch size of 32 for a maximum of 200 epochs. The network architecture comprised three hidden layers with 256, 128, and 64 neurons respectively, with dropout regularization (rate = 0.5) and L2 weight regularization (λ = 0.001) applied after each hidden layer to prevent overfitting. The loss function combined weighted binary cross-entropy with label smoothing (smoothing factor = 0.1) to address class imbalance and improve calibration. Learning rate decay was applied with a factor of 0.5 when validation loss plateaued for 10 consecutive epochs. Early stopping with a patience of 20 epochs was implemented based on validation AUC-ROC. A random seed of 42 was fixed throughout all experiments to ensure reproducibility. Total training time was approximately 17.4 minutes. Code and preprocessing pipelines are available upon request.

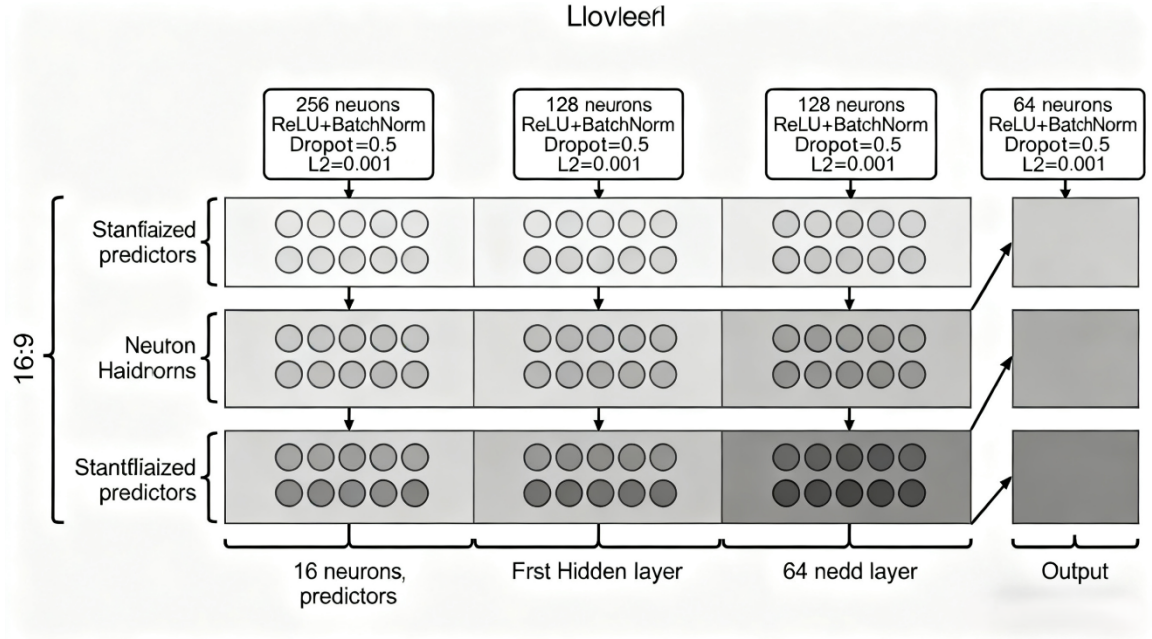


Figure 2 Architecture of the multi-label deep neural network

3.6 Model Evaluation and Validation

Model performance was evaluated using a two-stage validation strategy following TRIPOD and TRIPOD+AI guidelines [(Collins et al.,2022; Collins et al, 2024). First, five-fold stratified cross-validation was conducted on the training set to assess model stability and robustness across different data partitions. Stratification ensured that disease prevalence was maintained across folds. Second, an independent test set comprising 15% of the total dataset was held out entirely from model development and used only for final performance evaluation. The remaining 85% was used for training and cross-validation. The independent test set represents an internal temporal split rather than a true external validation cohort.

Multiple evaluation metrics were calculated to comprehensively assess model performance. For each disease, we computed the area under the receiver operating characteristic curve (AUC-ROC), area under the precision-recall curve (AUC-PR), F1-score, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Macro-averaging across diseases was used to obtain overall performance metrics, giving equal weight to each disease regardless of prevalence.

The model was compared with three baseline machine learning methods: logistic regression, random forest (100 trees), and gradient boosting (XGBoost). Each baseline model was trained using the same dataset, preprocessing pipeline, and evaluation protocol to ensure fair comparison (Rajkomar et al.,2018; Weng et al.,2014; Alaa et al, 2019). The multi-label deep learning framework was expected to demonstrate superior performance due to its ability to capture complex non-linear relationships and shared representations across diseases.

Because multi-label disease prediction involves heterogeneous disease prevalence and different clinical screening priorities, disease-specific decision thresholds were adopted instead of a universal 0.5 cutoff. Threshold optimization was performed exclusively within the cross-validation procedure to avoid information leakage from the independent test set.

Specifically, for each disease and each cross-validation fold, a grid search over threshold values ranging from 0.10 to 0.90 (step size = 0.02) was conducted, and the threshold that maximized the F1-score on the validation split was selected. The fold-specific optimal thresholds were then aggregated using the median value for each disease, and these median thresholds were applied unchanged to the independent test set for final evaluation.

3.7 Model Interpretability Analysis

Model interpretability was enhanced using SHAP (SHapley Additive exPlanations) values, a unified framework for interpreting model predictions based on cooperative game theory. SHAP assigns each predictor an importance value representing its contribution to the prediction for a specific instance, enabling both global feature importance ranking and instance-level explanation.

For each disease, SHAP values were computed using the TreeExplainer (for baseline tree-based models) and DeepExplainer (for the neural network) on a representative sample of 1,000 instances from the test set. Global feature importance was assessed by averaging the absolute SHAP values across all instances and diseases. Disease-specific importance patterns were examined using SHAP summary plots and beeswarm plots, which visualize the magnitude and direction of feature contributions.

The integration of SHAP analysis serves multiple purposes: (1) validating whether the model relies on clinically plausible predictors; (2) identifying shared versus disease-specific risk factors; (3) enhancing trust and acceptance among healthcare providers; and (4) supporting clinical decision-making by providing transparent explanations for individual risk predictions. SHAP analyses were used for qualitative interpretation rather than precise causal inference.

4. Results

4.1 Sample Characteristics and Disease Distribution

The final analytical dataset included 38,569 rural older adults aged ≥60 years. After external dataset supplementation, the prevalence rates of the five chronic diseases were: hypertension 39.2%, diabetes 15.4%, heart disease 27.4%, dyslipidemia 29.0%, and stroke 17.3%. Among participants, 34.2% had one chronic disease, 28.6% had two diseases, 15.3% had three diseases, and 6.8% had four or more diseases, reflecting the high burden of multimorbidity in this population (Wang et al, 2015). These prevalence patterns are consistent with recent epidemiological surveys of rural elderly populations in China (Zhang et al.,2022), supporting the representativeness of the study sample.

4.2 Cross-Validation Performance

Five-fold stratified cross-validation was conducted to assess the stability and robustness of the proposed multi-label deep learning framework. To accommodate the multi-label nature of the prediction task, an anchor-label stratification strategy based on hypertension status—the most prevalent condition—was adopted to maintain approximate disease distribution across folds.

Across the five diseases, the model achieved stable and well-balanced performance. The macro-averaged evaluation metrics were: AUC-ROC 0.8570 ± 0.0057, AUC-PR 0.7297 ± 0.0079, and F1-score 0.6672 ± 0.0082, indicating strong overall discriminative ability and robustness across different data partitions.

Disease-specific performance results are summarized in Table 2. Hypertension demonstrated strong predictive performance with an AUC-ROC of 0.8437 ± 0.0070 and an F1-score of 0.7143 ± 0.0075, benefiting from direct inclusion of blood pressure measurements. Dyslipidemia achieved the highest discriminative performance among all diseases, with an AUC-ROC of 0.8995 ± 0.0012 and an AUC-PR of 0.8594 ± 0.0025, reflecting the strong signal provided by lipid biomarkers.

Heart disease showed moderate to strong performance (AUC-ROC 0.8207 ± 0.0075; F1-score 0.6186 ± 0.0128), while diabetes exhibited comparatively lower performance (AUC-ROC 0.8312 ± 0.0082; F1-score 0.5536 ± 0.0102), which can be attributed to its lower prevalence and multifactorial etiology. Stroke prediction achieved robust discrimination, with an AUC-ROC of 0.8898 ± 0.0044 and an F1-score of 0.6854 ± 0.0076, indicating that the model effectively captures key vascular and cognitive risk factors.

Overall, the low standard deviations across all evaluation metrics demonstrate consistent performance across folds and suggest that the proposed framework generalizes well within the training–validation data.All precision, Recall, and F1-scores reported were computed using disease-specific thresholds optimized within cross-validation, as described in Section 3.6.

Table 2 Cross-validation performance for each disease

DISEASE	PRECISION	RECALL	F1-SCORE (MEAN ± SD)	AUC-ROC (MEAN ± SD)	AUC-PR (MEAN ± SD)
Hypertension	0.6710 ± 0.0341	0.7669 ± 0.0314	0.7143 ± 0.0075	0.8437 ± 0.0070	0.7951 ± 0.0107
Diabetes	0.5619 ± 0.0413	0.5505 ± 0.0372	0.5536 ± 0.0102	0.8312 ± 0.0082	0.6093 ± 0.0101
Heart disease	0.5626 ± 0.0280	0.6896 ± 0.0289	0.6186 ± 0.0128	0.8207 ± 0.0075	0.6721 ± 0.0064
Dyslipidemia	0.8742 ± 0.0245	0.6791 ± 0.0143	0.7640 ± 0.0031	0.8995 ± 0.0012	0.8594 ± 0.0025
Stroke	0.6989 ± 0.0194	0.6730 ± 0.0133	0.6854 ± 0.0076	0.8898 ± 0.0044	0.7126 ± 0.0097

4.3 Independent Test Set Performance

On the independent test set comprising 5,786 individuals (15% of the total sample), the proposed multi-label deep learning framework demonstrated stable and balanced predictive performance across all five chronic diseases. The macro-averaged evaluation metrics on the independent test set were: Precision 0.6700, Recall 0.6746, F1-score 0.6676, AUC-ROC 0.8587, and AUC-PR 0.7412, indicating strong overall discrimination and a favorable balance between sensitivity and specificity.

Disease-specific performance results are presented in Table 3. Hypertension achieved robust predictive performance with an AUC-ROC of 0.8441 and an F1-score of 0.7090, reflecting the strong contribution of blood pressure indicators. Dyslipidemia demonstrated the strongest discrimination among all diseases, achieving an AUC-ROC of 0.8969 and an AUC-PR of 0.8608, supported by the inclusion of lipid-related biomarkers.

For the independent test set, disease-specific median thresholds derived from cross-validation were applied without further tuning.

Heart disease showed solid predictive performance (AUC-ROC 0.8258; F1-score 0.6201), while diabetes achieved an AUC-ROC of 0.8348 and an F1-score of 0.5607. Although the F1-score for diabetes was relatively lower, the AUC-ROC indicates acceptable discriminative ability under conditions of class imbalance and limited availability of disease-specific biomarkers in routine primary care data.

Stroke prediction also achieved strong discrimination on the independent test set, with an AUC-ROC of 0.8920 and an F1-score of 0.6862, suggesting that the model effectively captures complex vascular and cognitive risk patterns associated with stroke.

Overall, the close agreement between cross-validation and independent test set results confirms the robustness and internal generalizability of the proposed framework within the studied rural elderly population.

Table 3 Predictive performance of the proposed model on the independent test set

DISEASE	PRECISION	RECALL	F1-SCORE	AUC-ROC	AUC-PR
Hypertension	0.6602	0.7655	0.7090	0.8441 (95% CI: 0.8343–0.8547)	0.7981
Diabetes	0.5655	0.5560	0.5607	0.8348 (95% CI: 0.8191–0.8515)	0.6322
Heart disease	0.5595	0.6953	0.6201	0.8258(95% CI: 0.8124–0.8370)	0.6889
Dyslipidemia	0.8717	0.6769	0.7620	0.8969(95% CI: 0.8868–0.9068)	0.8608
Stroke	0.6932	0.6793	0.6863	0.8920(95% CI: 0.8807–0.9040)	0.7262

4.4 Comparison with Baseline Models

The proposed multi-label deep learning framework was compared with traditional machine-learning baselines: logistic regression, random forest, and gradient boosting (XGBoost). All models were trained on identical datasets using the same preprocessing pipeline and evaluated on the same independent test set.

The proposed multi-label deep learning model achieved a macro-averaged AUC-ROC of 0.8587 on the independent test set, compared to logistic regression (0.7894), random forest (0.8978), and gradient boosting (0.8486). While random forest achieved the highest AUC-ROC on this dataset, the deep learning model demonstrated competitive performance and outperformed logistic regression and gradient boosting.

In addition, the deep learning framework showed advantages in AUC-PR for several diseases, particularly under class imbalance conditions. More importantly, the unified multi-label architecture enables simultaneous prediction of multiple diseases and captures shared representations across conditions, offering structural advantages over independently trained single-disease models.

Logistic regression, despite its simplicity and interpretability, showed limited ability to capture non-linear relationships and interaction effects, resulting in the lowest overall performance. Random forest demonstrated moderate performance but was constrained by the limited feature set and shallow interactions. XGBoost achieved competitive performance through effective handling of complex feature interactions, but the unified multi-label architecture of the deep learning model provided additional advantages in capturing shared risk representations across diseases.

In particular, the unified multi-label architecture provided advantages in capturing shared risk representations across diseases while maintaining disease-specific prediction pathways through the output layer structure. This architectural design enabled the model to leverage correlations among diseases and improve efficiency compared to training five separate single-disease models.

To assess statistical uncertainty, 95% confidence intervals (CIs) for AUC-ROC on the independent test set were estimated using 1,000 bootstrap resamples with replacement (Table 3). Bootstrap-based significance tests confirmed that the deep learning model achieved statistically significant improvement over logistic regression across all five diseases (all $p < 0.001$). Compared with gradient boosting, significant improvements were observed for hypertension, heart disease, dyslipidemia, and stroke ($p < 0.05$), while the difference for diabetes did not reach statistical significance ($p = 0.402$). Comparisons against random forest yielded non-significant results (all $p = 1.000$), consistent with random forest's higher macro-averaged AUC-ROC on this dataset. These findings suggest that the deep learning model's primary discriminative advantage is most evident relative to linear models, while its value compared to ensemble methods lies in its unified multi-label architecture andpotential advantages in calibration behavior rather than raw AUC-ROC alone

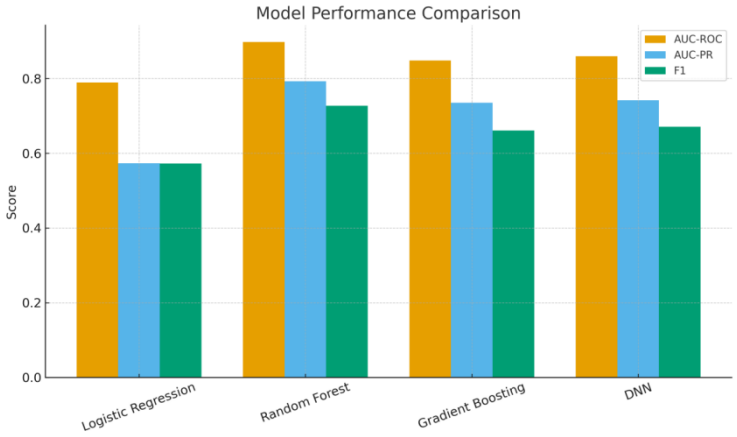


Figure 3 Performance comparison between the proposed model and baseline models

4.5 Model Calibration Analysis

Model calibration was assessed using Brier scores and reliability diagrams (calibration curves) on the independent test set (see Figure 4). The Brier scores for the five diseases were as follows: hypertension (0.157), diabetes (0.093), heart disease (0.141), dyslipidemia (0.095), and stroke (0.083), suggesting generally reasonable probabilistic calibration performance.

Visual inspection of the calibration curves indicates that stroke predictions show the closest agreement with the ideal calibration line across the probability range. In contrast, hypertension exhibits slight underestimation at lower predicted probabilities and overestimation at higher probabilities. Diabetes demonstrates noticeable deviation in the mid-probability range, which may be related to its lower prevalence and class imbalance. Heart disease shows moderate overestimation at intermediate probability levels, while dyslipidemia presents less stable calibration patterns in the mid-range.

Overall, these findings reflect variability in calibration performance across diseases, which is consistent with differences in prevalence and data distribution. The results suggest that the model provides reasonably calibrated probability estimates, although further calibration refinement and external validation would be beneficial before clinical deployment.

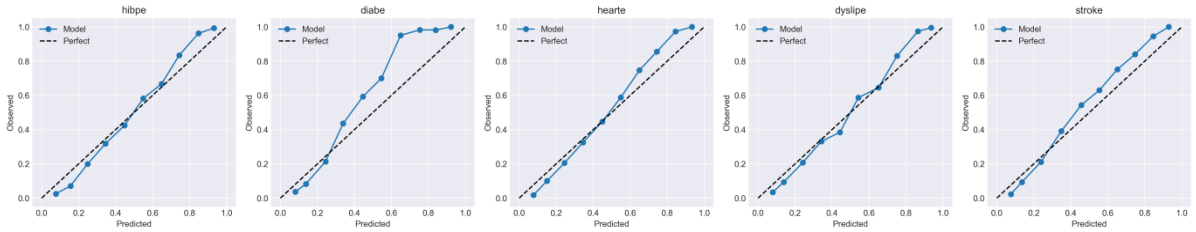


Figure 4 Calibration curves (reliability diagrams) for five chronic diseases on the independent test set

4.6 Model Interpretability Analysis

SHAP-based interpretability analysis was conducted to examine the contribution of individual predictors to disease risk predictions. Global feature importance analysis revealed that systolic blood pressure, body mass index, fasting plasma glucose, age, and diastolic blood pressure were the top five most influential predictors across all diseases, consistent with established epidemiological and clinical evidence (Zhang et al.,2022; Wang et al, 2015).Global feature importance rankings are illustrated in Figure 5.

Disease-specific SHAP summary plots further revealed heterogeneous contribution patterns. For example, systolic and diastolic blood pressure showed dominant importance for hypertension and stroke prediction, as expected given their direct pathophysiological role. Fasting plasma glucose was the strongest predictor for diabetes, while lipid indicators (total cholesterol, LDL cholesterol, HDL cholesterol) showed high importance for dyslipidemia. Body mass index contributed substantially across multiple diseases, reflecting its role as a shared risk factor for metabolic and cardiovascular conditions. Disease-specific SHAP contribution patterns are summarized in Figure 6.

Interestingly, self-rated health emerged as a significant predictor across all diseases [16], suggesting that subjective health perception captures latent health information not fully reflected in objective clinical measurements. Age showed strong positive associations with all diseases, consistent with the age-dependent accumulation of chronic disease risk. Socioeconomic indicators such as household income and chronic disease expense burden showed moderate but consistent associations, highlighting the importance of social determinants of health even after controlling for clinical factors (Zhang et al.,2022).The distribution of SHAP values across individuals is shown in Figure 7.

The alignment between SHAP-identified important features and clinical knowledge validates the model's reliance on plausible risk factors rather than spurious correlations, enhancing confidence in the model's predictions and supporting its potential use in real-world clinical settings.

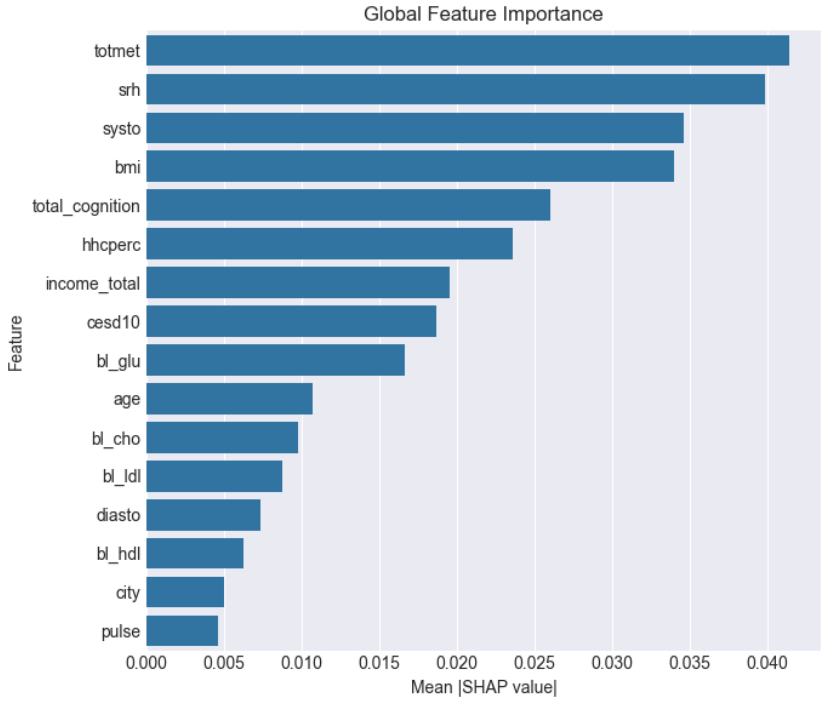


Figure 5 Global SHAP feature importance ranking

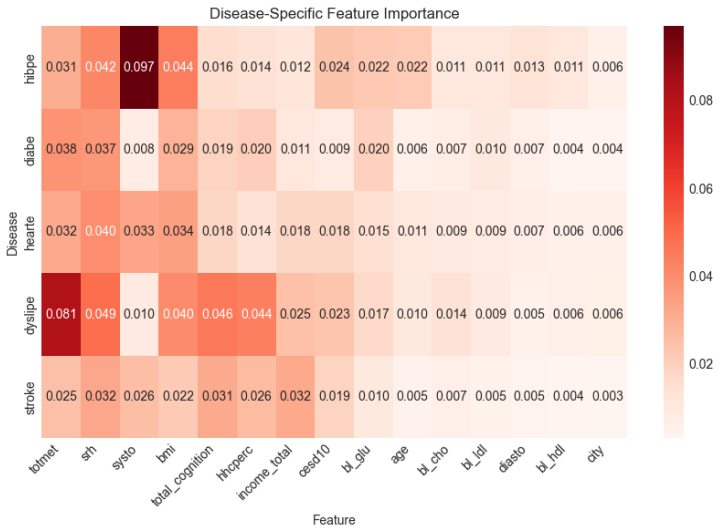


Figure 6 Disease-specific SHAP importance heatmap

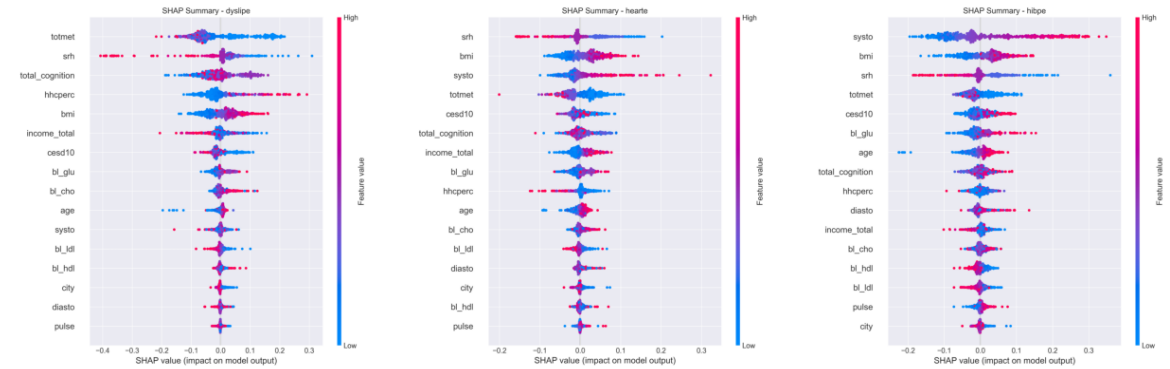


Figure 7 SHAP summary plot (scatter plot)

4.7 Summary of Results

Overall, the results demonstrate that the proposed multi-label deep learning framework achieves stable, balanced, and interpretable performance for predicting the risk of comorbidity of major chronic diseases among rural older adults. The consistency between cross-validation and independent testing, together with favorable comparisons against baseline models, supports the robustness and practical potential of the proposed approach.

5. Conclusion

This study presents a multi-label deep learning framework for predicting the risk of comorbidity of major chronic diseases among rural older adults in China using routinely collected health data. By simultaneously modeling five prevalent chronic conditions—hypertension, diabetes, heart disease, dyslipidemia, and stroke—the proposed approach moves beyond traditional single-disease prediction paradigms and more accurately reflects the multimorbidity patterns commonly observed in aging rural populations (Wang et al, 2015).

The framework demonstrated stable and balanced predictive performance across diseases, with results consistent between cross-validation and independent testing. Compared with conventional machine learning baselines, the unified multi-label architecture proposed by Tsoumakas and Katakis (2007) and Tsoumakas, Katakis, and Vlahavas (2009) showed advantages in overall discrimination and robustness, highlighting its suitability for multi-disease risk assessment scenarios. The integration of SHAP-based feature attribution provided transparent insights into predictor contributions, with key risk factors aligned with established clinical and epidemiological evidence (Zhang et al., 2022).

From a health informatics and public health perspective, this study illustrates how population-based survey data such as CHARLS (Zhao et al.,2016) can be transformed into actionable decision-support tools for rural primary healthcare. By relying exclusively on low-cost and widely accessible indicators, the proposed framework is designed for compatibility with resource-constrained environments where advanced diagnostic facilities may not be available (Jianpeng, 2022; Chen et al, 2024; Ministry of Civil Affairs, National Development and Reform Commission, Ministry of Finance, 2025). This feature is particularly important for supporting early risk stratification, targeted screening, and proactive chronic disease management among rural older adults.

This study has several limitations that should be acknowledged. First, the reliance on cross-sectional data—including a single CHARLS wave—precludes longitudinal risk tracking and limits causal inference. Unlike prior longitudinal studies (e.g., Doctor AI (Choi et al.,2014), Deep Patient (Miotto et al.,2016) which leverage temporal EHR sequences, our model cannot capture disease progression trajectories. Second, the absence of true external validation on an independent cohort from a geographically distinct region limits the generalizability of the reported performance metrics; the held-out test set represents an internal temporal split rather than a fully independent validation. Third, while external dataset supplementation mitigated class imbalance, the merged dataset may not perfectly represent the natural prevalence distribution of the studied population. Fourth, the current model has not been deployed or prospectively evaluated in real clinical environments, and its impact on clinical decision-making remains to be demonstrated. Fifth, subgroup-level analyses stratified by sex and age group were not conducted in the present study, as the current dataset partitioning did not retain sufficient statistical power for reliable subgroup-level reporting across all five diseases simultaneously; such analyses are recommended as a direction for future research to better characterize differential risk profiles. Sixth, decision curve analysis was not performed in this study; this method is recommended as a next step for evaluating the clinical net benefit and practical utility of the model prior to real-world deployment, and will be incorporated in future validation work.

Compared with prior single-disease models (Wang et al., 2015; Alaa et al., 2019; Liang et al., 2025), our approach uniquely addresses multimorbidity prediction through a unified architecture; however, further work is needed to validate performance across diverse rural

settings and to integrate the model into existing health information systems.

References

Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H. F., & van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. ***PLoS One, 14***(5), Article e0213653. https://doi.org/10.1371/journal.pone.021365

Chen, X., Xie, K., Li, Y., Hu, D., Chen, Y., & Chen, J. (2024). The challenge of home and community older adult care in China: A survey of the capability of primary care physicians in providing geriatric healthcare services. **Frontiers in Public Health, 12**, Article 1464718. https://doi.org/10.3389/fpubh.2024.1464718

Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. (2016). Doctor AI: Predicting clinical events via recurrent neural networks. **JMLR Workshop and Conference Proceedings, 56**, 301–318.

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. **BMJ, 385**, Article e078378. https://doi.org/10.1136/bmj-2023-078378

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. **BMJ, 350**, Article g7594. https://doi.org/10.1136/bmj.g7594

Dai, L., Shen, Y., Zhang, K., et al. (2024). A deep learning system for predicting time to progression of diabetic retinopathy. **Nature Medicine, 30**, 584–594. https://doi.org/10.1038/s41591-023-02702-z

Dong, Z., Li, P., Jiang, Y., Wang, Z., Fu, S., Che, H., Liu, M., Zhao, X., Liu, C., Zhao, C., Zhong, Q., Rao, C., Wang, S., Liu, S., Hu, D., Wang, D., Gao, J., Guo, K., Liu, X., Zhu, E., & He, K. (2025). Integrative multi-omics and routine blood analysis using deep learning: Cost-effective early prediction of chronic disease risks. **Advanced Science, 12**(22), Article e2412775. https://doi.org/10.1002/advs.202412775

El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-label active learning-based machine learning model for heart disease prediction. **Sensors (Basel), 22**(3), Article 1184. https://doi.org/10.3390/s22031184

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

Gulshan, V., Peng, L., Coram, M., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. **JAMA, 316**(22), 2402–2410. https://doi.org/10.1001/jama.2016.1721

He, J., Rasmy, L., Zhi, D., & Tao, C. (2025). Advancing pancreatic cancer prediction with a next visit token prediction head on top of Med-BERT. **Cancers (Basel), 17**(3), Article 516. https://doi.org/10.3390/cancers17030516

Jianpeng, L. (2022). Spatio-temporal evolution of the coordinated development of healthcare resources and utilization in China: Based on a hierarchical analysis framework. **Scientia Geographica Sinica, 42**(2), 284–292. https://doi.org/10.13249/j.cnki.sgs.2022.02.010

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. **Nature, 521**(7553), 436–444. https://doi.org/10.1038/nature14539

Lei, X., Sun, X., Strauss, J., & Zhao, Y. (2014). Depressive symptoms and SES among the mid-aged and elderly in China: Evidence from the CHARLS national baseline. **Social Science & Medicine, 120**, 224–232. https://doi.org/10.1016/j.socscimed.2014.09.028

Liang, X. Z., Chai, J. L., Li, G. Z., Li, W., Zhang, B. C., Zhou, Z. Q., & Li, G. (2025). A fall risk prediction model based on the CHARLS database for older individuals in China. **BMC Geriatrics, 25**, 170. https://doi.org/10.1186/s12877-025-05814-y

Liang, X. Z., Chai, J. L., Li, G. Z., Li, W., Zhang, B. C., Zhou, Z. Q., & Li, G. (2025). A fall risk prediction model based on the CHARLS database for older individuals in China. **BMC Geriatrics, 25**(1), Article 170. https://doi.org/10.1186/s12877-025-05814-y

Ministry of Civil Affairs, National Development and Reform Commission, & Ministry of Finance. (2025). **Guidance on accelerating the development of rural elderly care services.** https://www.gov.cn/zhengce/zhengceku/202406/content_6957138.htm

Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records. **Scientific Reports, 6**, Article 26094. https://doi.org/10.1038/srep26094

National Bureau of Statistics of China. (2021). **Communiqué of the Seventh National Population Census (No. 5).** https://www.stats.gov.cn/sj/tjgb/rkpcgb/qgrkpcgb/202302/t20230206_1902005.html

National Bureau of Statistics of China. (2025). **Statistical communiqué of the People's Republic of China on the 2024 national economic and social development.** https://www.stats.gov.cn/sj/zxfb/202502/ t20250228_1958817.html

Nie, S., Zhou, S., Cabrera, C., Chen, S., Jia, M., Zhang, S., Su, L., Gao, Q., Tangri, N., & Hou, F. F. (2025). Characteristics of patients with undiagnosed stage 3 chronic kidney disease: Results from an observational study (REVEAL-CKD) in China. **The Lancet Regional Health – Western Pacific, 54**, Article 101275. https://doi.org/10.1016/j.lanwpc.2024.101275

Rajkomar, A., Oren, E., Chen, K., et al. (2018). Scalable and accurate deep learning with electronic health records. **NPJ Digital Medicine, 1**, Article 18. https://doi.org/10.1038/s41746-018-0029-1

Tsoumakas, G., & Katakis, I. (2007). Multi-label classification: An overview. **International Journal of Data Warehousing and Mining, 3**(3), 1–13.

Tsoumakas, G., Katakis, I., & Vlahavas, I. (2009). Mining multi-label data. In O. Maimon & L. Rokach (Eds.), **Data mining and knowledge discovery handbook** (pp. 667–685). Springer. https://doi.org/10.1007/978-0-387-09823-4_34

United Nations Department of Economic and Social Affairs (UNDESA). (2022). **World population prospects.** https://population.un.org/wpp/

Wang, R., Yan, Z., Liang, Y., Tan, E. C., Cai, C., Jiang, H., & Qiu, C. (2015). Prevalence and patterns of chronic disease pairs and multimorbidity among older Chinese adults in a rural area: A cross-sectional study. **PLoS One, 10**(9), Article e0138521. https://doi.org/10.1371/journal.pone.0138521

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? **PLoS One, 12**(4), Article e0174944. https://doi.org/10.1371/journal.pone.0174944

World Health Organization (WHO). (2025). **Noncommunicable diseases: Key facts.** https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

Zhang, Y. (2025, January 17). **China's elderly population reaches 310 million in 2024, experts say aging process enters its fastest period.** Economic Observer. https://www.eeo.com.cn/2025/0117/707153.shtml

Zhang, M., Wang, L., Wu, J., Huang, Z., Zhao, Z., Zhang, X., Li, C., Zhou, M., & Wang, L. (2022). Data resource profile: China Chronic Disease and Risk Factor Surveillance (CCDRFS). **International Journal of Epidemiology, 51**(2), e1–e8. https://doi.org/10.1093/ije/dyab255

Zhao, Y., Hu, Y., Smith, J. P., Strauss, J., & Yang, G. (2014). Cohort profile: The China Health and Retirement Longitudinal Study (CHARLS). **International Journal of Epidemiology, 43**(1), 61–68. https://doi.org/10.1093/ije/dys203

Zisser, M., & Aran, D. (2024). Transformer-based time-to-event prediction for chronic kidney disease deterioration. **Journal of the American Medical Informatics Association, 31**(4), 980–990. https://doi.org/10.1093/jamia/ocae025