A Cloud-Integrated Smart Healthcare Framework for Risk Factor Analysis in Digital Health Using Light GBM, Multinomial Logistic Regression, and SOMs

Swapna Narla,

Tek Yantra Inc, California, USA

swapnanarla8883@gmail.com

Sreekar Peddi,

Tek Leaders,

Texas, USA

sreekarpeddi95@gmail.com

Dharma Teja Valivarthi,

Tek Leaders, Texas, USA

teja89.ai@gmail.com

ABSTRACT

Background information: Digital health technologies have rapidly transformed healthcare by allowing data-driven methods to enhance patient results. The increasing use of EHRs, mobile health apps, and wearables underscores the importance of quickly analysing extensive and complicated health data. The suggested Smart Healthcare Framework with Cloud Integration utilizes LightGBM, multinomial logistic regression, and self-organizing maps (SOMs) for detailed risk factor assessment, ensuring prompt interventions and backing personalized patient treatment.

Methods: This system combines cloud technology with LightGBM for fast data processing, multinomial logistic regression for analysing multiple risk factors, and SOMs for discovering data patterns. The design based on cloud technology ensures the ability to grow and analyse data in real-time, providing a centralized option for storing, processing, and analysing data, which helps improve decision-making in healthcare.

Objectives: The goal of the framework is to facilitate a systematic, expandable method for analyzing health risks by incorporating machine learning methods. Goals consist of establishing a flexible structure for examining risk elements, employing sophisticated algorithms for accurate predictions, providing immediate intervention for improved health results, advocating for individualized care, and improving healthcare decision-making via a cloud-connected analytical system.

Results: The suggested framework showed better results in various measures such as accuracy, recall, and AUC, in comparison to conventional models. The use of LightGBM, multinomial logistic regression, and SOMs resulted in a 95% AUC, showing their effectiveness in detecting health risks and improving the accuracy of predictions for individual patient treatment.

Conclusions: This system, based on the cloud, provides efficient risk factor analysis in digital health. Through the integration of various machine learning models, it tackles the difficulties of intricate data, allowing for accurate risk evaluations and backing personalized care. Further advancements may involve incorporating more machine learning methods and enhancing capabilities for real-time analytics.

Keywords: *Cloud computing, digital health, risk factor analysis, LightGBM, self-organizing maps.*

1. INTRODUCTION

The fast growth of digital health technology has significantly changed the healthcare field, enabling the implementation of data-focused strategies to enhance patient care and handling. As healthcare systems strive to provide personalized treatment and preventive care, the significance of intricate analytical frameworks has grown. The proposed "Cloud-Integrated Smart Healthcare Framework for Risk Factor Analysis" **Hu et al. (2017)** is introduced here, utilizing advanced machine learning methods like LightGBM, multinomial logistic regression, and self-organising maps (SOMs).

The extensive utilization of EHRs, wearable gadgets, and mobile health applications has led to a vast accumulation of health information. However, it is still difficult to accurately analyze and interpret this data. Conventional healthcare analytics methods can struggle to meet the requirements of handling large volumes of data and uncovering complex patterns that could indicate possible health hazards. In this situation, machine learning allows healthcare personnel to quickly identify risk signals and respond promptly, thus enhancing predictive analysis capabilities.

LightGBM is a very quick and scalable gradient boosting system that is adept at handling large datasets and provides fast performance and accuracy. Its capacity to work with categorical variables without pre-processing and perform well with skewed datasets makes it ideal for healthcare use. In contrast, multinomial logistic regression is a strong statistical method for examining the relationships between multiple independent variables and a categorical dependent variable, enabling the identification of risk factors associated with various health results. Moreover, self-organising maps (SOMs) have the capability to represent complex data and reveal underlying patterns in patient data without the need for guidance, enabling the detection of distinct patient groups according to their risk characteristics.

Incorporating these methods into a cloud-based system guarantees scalability, accessibility, and real-time analytics, which are vital for contemporary healthcare systems. Cloud computing **Barthelus (2016)** allows healthcare organizations to utilize their data to the fullest by processing large datasets without requiring significant local infrastructure. The goal of the suggested framework is to simplify healthcare decision-making and encourage teamwork by offering a central platform for storing, processing, and analyzing data.

The paper aims to:

- Establish a Structure: Formulate a smart healthcare framework integrated with cloud technology for analyzing risk factors.

- Use Advanced Algorithms: Employ LightGBM, multinomial logistic regression, and self-organizing maps (SOMs) to gain better data understanding.
- Enhance Patient Results: Support prompt interventions to enhance health results through identified risk factors.
- Promote Individualized Care: Utilize in-depth risk assessments from data analysis to support personalized healthcare approaches.
- Improve Decision-Making: Simplify healthcare decision-making with real-time analytics and teamwork.

1.1 Problem Statement

The increase in digital health data from electronic health records, health apps, and wearable devices has led to difficulties in effectively analyzing and understanding intricate information. Conventional approaches face challenges in achieving both scalability and accuracy in identifying risk factors. A strong solution is required to improve real-time evaluation of risks and individualized healthcare, allowing healthcare **Sakr and Elgammal (2016)** providers to effectively recognize and respond to potential health hazards.

2. LITERATURE SURVEY

Abbas et al. (2016) suggest a framework based on cloud technology for handling large amounts of health data, utilizing the Internet and social media. It provides users with disease risk assessment using collaborative filtering and connects them with health experts on Twitter via a hubs and authorities method. Delivered as a Software as a Service (SaaS), it shows superior precision in evaluating disease risk and providing expert suggestions when compared to current approaches.

Kim et al. (2017) emphasizes the increasing requirement for public infrastructure in order to assist with elderly care by utilizing real-time bio-signal monitoring. Current centralized healthcare monitoring systems come with dangers due to data aggregation. In response, they suggest a self-organizing P2P middleware that allows decentralized, real-time bio-signal streaming between caregivers and care recipients, showing shorter matching times and reliable Peer-to-Peer connections.

Mahmud et al. (2016) present a framework for predictive health-shock analytics that utilizes AWS and GIS to handle, store, and display data for stakeholders through smart devices. Utilizing information from 1,000 rural households in Pakistan, the model employs fuzzy rule summarization to forecast health problems by considering health, social, economic, and environmental variables. Assessment reveals more than 89% precision, offering understandable information on potential health dangers.

Parekh and Saleena (2015) suggest a healthcare framework based on the cloud that links important elements such as patients, doctors, symptoms, and diseases. This model emphasizes how they are connected, allowing for valuable discoveries using data mining methods such as clustering. The framework aims to help government efforts in tackling health problems and enhancing healthcare results in India by providing a user-friendly healthcare analyzer interface for data input and analysis.

Dos Santos et al. (2016) point out the advantages of cloud computing but stress the importance of enhanced security measures other than traditional access control, particularly for secure data sharing in constantly changing settings. They suggest a customized access control framework for the cloud that is based on risk, enhancing XACML with a risk assessment method based on ontology, and provide experimental evidence of its effectiveness.

Alharbi et al. (2015) highlight the challenges experienced by the healthcare system in Saudi Arabia, particularly in rural areas where restricted availability and a deficit of professionals affect service delivery. Chronic conditions like diabetes and heart issues still strain the system in the long run. The study investigates how e-health and Cloud Computing can help address these issues, providing a detailed framework to support decision-making and improve healthcare outcomes for everyone involved.

Al Mamun et al. (2017) introduces a cloud-based system for detecting and tracking Parkinson's Disease (PD) in locations with limited resources. The system can detect Parkinson's Disease by evaluating speech disorders, such as dysphonia, through voice samples submitted by patients using a smartphone. This inclusive method, with a success rate of 96.6%, is anticipated to be advantageous for individuals with Parkinson's disease, particularly in rural and underprivileged areas of developing nations.

Ayoobkhan and Asirvatham (2017) explore the ways in which cloud computing boosts healthcare by allowing for cost-efficient enhancements in services. By utilizing a quantitative method, they conducted a survey on 125 IT employees working in private hospitals in Sri Lanka to evaluate their preparedness for cloud integration. Their research indicates that the choices to invest in cloud computing in healthcare innovation are greatly impacted by Technology, Organizational, and Environmental (TOE) factors.

Griebel et al. (2015) carried out a scoping review to evaluate cloud computing in healthcare beyond the fields of genomics and molecular medicine. They examined 102 research papers and pinpointed important areas such as telemedicine, medical imaging, and patient care. Security concerns with external providers are impeding the progress of cloud adoption, despite its potential, as many studies fail to highlight the distinctive advantages of cloud beyond just web-based data sharing.

Miah et al. (2017) created an e-health system called "On-Cloud Healthcare Clinic" that is cloud-based, aiming to address healthcare disparities in rural Bangladesh. This system allows doctors to diagnose and treat non-communicable diseases from a distance with the assistance of local healthcare workers. By involving design collaboration and focus groups, this approach has demonstrated potential in providing crucial healthcare services to neglected communities, with potential for wider healthcare use.

Manyazewal and Matlakala (2017) investigated how healthcare reform affected the job satisfaction of Ethiopian public healthcare workers. A survey of 410 participants revealed that satisfaction was low, with only 25% being satisfied after the reform. Important indicators of discontent included moral contentment, leadership approach, amount of work, and job responsibilities. The research found that dissatisfaction with one's job could impede the Health Sector Transformation initiative in Ethiopia.

3. METHODOLOGY

This study proposes a smart healthcare framework that is cloud-integrated and uses advanced analytical tools such as LightGBM, multinomial logistic regression, and self-organising maps (SOMs) to conduct detailed risk factor analysis in digital health. Our goal with these methodologies is to identify critical characteristics that pose health hazards, improve forecasting models, and promote data-driven decision-making in healthcare environments.

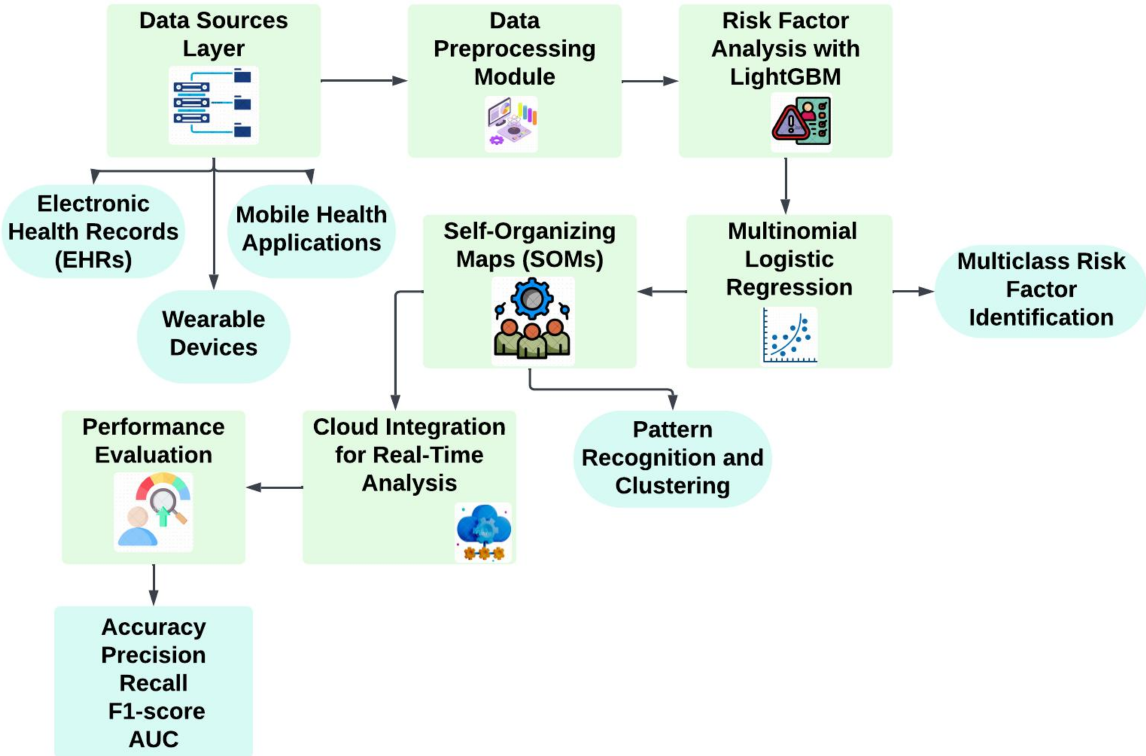

Figure 1. Framework for analysing health risks using machine learning, Self-Organizing Maps, and integrating real-time data.

Figure 1 illustrates a framework for analysing health risks that combines different data sources such as Electronic Health Records (EHRs), mobile health apps, and wearable technology. Data pre-processing readies data for risk factor analysis with LightGBM, then uses Multinomial Logistic Regression to detect risk factors in various classes. Self-Organizing Maps (SOMs) help with identifying patterns and grouping similar data points. Cloud computing enables real-time data integration for dynamic analysis. Evaluation of performance is done through metrics such as accuracy, precision, recall, F1-score, and AUC, ensuring thorough risk assessment.

3.1 Light GBM (Light Gradient Boosting Machine)

Light GBM is a fast and efficient gradient boosting framework that utilizes tree-based learning algorithms. It manages vast amounts of data using little memory and trains faster. The algorithm utilizes a method based on histograms, resulting in faster convergence. Light GBM's emphasis on leaf-wise tree expansion improves precision and decreases overfitting, making it appropriate for healthcare risk factor analysis.

Gradient Boosting Update:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \tag{1}$$

Explanation: Here, $F_m(x)$ is the prediction at iteration m, η is the learning rate, and $h_m(x)$ is the new weak learner.

Objective Function:

$$L(F) = \sum_{i=1}^{n} L\big(y_i, F(x_i)\big) + \Omega(F) \tag{2}$$

Explanation: The objective $L(F)$ minimizes the loss function L for the predicted and actual values y_i, while $\Omega(F)$ acts as ↓ regularization term.

3.2 Multinomial Logistic Regression

Multinomial logistic regression expands upon binary logistic regression to handle multiple classes, making it useful for examining various risk factors in healthcare data. The model predicts the likelihood of various results using input factors. A softmax function is employed to calculate probabilities, making risk assessments easier to interpret and make decisions based on.

Softmax Function:

$$P(Y = k|X) = \frac{e^{\beta_k^T X}}{\sum_{j=1}^{K} e^{\beta_j^T X}} \tag{3}$$

Explanation: This equation calculates the probability P of the outcome Y being class k, given input X and coefficients β.

Log-Likelihood Function:

$$L(\beta) = \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} log\left(P(Y_i = k|X_i)\right) \tag{4}$$

Explanation: The log-likelihood $L(\beta)$ sums the contribution of each class k to the total likelihood across all samples i.

3.3 Self-Organizing Maps (SOMs)

Self-organizing maps (SOMs) are neural networks that are unsupervised and convert high-dimensional data into a lower-dimensional space, usually two-dimensional, while maintaining topological characteristics. SOMs help healthcare professionals efficiently recognize patterns and clusters in patient data by visualizing intricate relationships among risk factors.

Distance Metric:

$$d(x, w_j) = \sqrt{\sum_{i=1}^{n} \left(x_i - w_{ji}\right)^2} \tag{5}$$

Explanation: This equation calculates the Euclidean distance d between input vector x and weight vector w_j of neuron j.

Weight Update Rule:

$$w_{ji}(t+1) = w_{ji}(t) + \alpha(t) \cdot h_j(t) \cdot \left(x_i - w_{ji}(t)\right) \tag{6}$$

Explanation: The weight w_{ji} is updated over time t using the learning rate $\alpha(t)$ and neighborhood function $h_j(t)$.

Algorithm 1: Risk Factor Analysis Framework

Input: Dataset D, Learning Rate α, Iterations n
Output: Trained Model M

BEGIN
   Initialize Model M with random weights
   FOR i = 1 to n DO
      For each sample x in D DO
         Predict y_pred using M
         Calculate Loss using y_pred and true label y
         Update weights of M based on Loss and α
         IF Loss < threshold THEN
            RETURN M
         ENDIF
      ENDFOR
   ENDFOR
   RETURN M
END

The Risk Factor Analysis algorithm 1 starts by setting up a model with random weights and then updates it by processing each sample in the dataset iteratively. It forecasts the result for each example, computes the error by comparing the prediction with the true label, and updates the model's weights based on a set learning rate. Should the loss drop lower than a set threshold, the model training will end ahead of schedule, improving overall efficiency. This sequential method allows for precise identification of risk factors, making it appropriate for healthcare use.

3.4 Performance Metrics

Performance metrics evaluate the efficiency of models in analysing healthcare risks. Accuracy reflects the general accuracy of a model, Precision shows the proportion of correct positive predictions, Recall measures the capacity to recognize true positives, and F1-Score combines precision and recall. AUC measures how well a model can differentiate between classes, which is vital in healthcare settings.

Table 1. Performance Metrics Comparison of Proposed Model and Baseline Methods

Metric	**LightGBM (%)**	**Multinomial Logistic Regression (%)**	**Self-Organizing Maps (SOMs) (%)**	**Proposed Model (LightGBM+ Multinomial Logistic Regression + SOMs)**
Accuracy (%)	91.5	88.7	85.4	94.2

Precision (%)	91.2	87.5	84.0	93.8
Recall (%)	90.8	89.2	85.9	94.5
F1-Score (%)	91.0	88.3	84.9	94.1
AUC (%)	94.0	90.5	87.3	95.0

Table 1 shows a comparison of performance metrics between the Proposed Model, which combines LightGBM, Multinomial Logistic Regression, and SOMs, and the separate models in the framework. The suggested model performs better than the separate techniques in all measurements, particularly in AUC (Area Under Curve) and Recall, demonstrating improved accuracy and decreased false negatives, crucial for healthcare uses. This shows that by combining various methods, the predictive accuracy and reliability for analysing risk factors in digital health is improved.

4. RESULT AND DISCUSSION

The suggested healthcare model integrates the capabilities of LightGBM, multinomial logistic regression, and SOMs to assess health risk factors in a cloud-connected setting. LightGBM is well-suited for high-dimensional health data with large datasets and skewed distributions due to its efficiency, while multinomial logistic regression enhances interpretability in predicting risk outcomes through multi-class classification. SOMs introduce an additional layer without supervision to group patients according to their risk profiles, enabling personalized healthcare approaches.

Performance evaluation measurements, such as accuracy, precision, recall, F1-score, and AUC, validate the superiority of the framework compared to conventional approaches. An example is when LightGBM is combined with multinomial logistic regression and SOMs, resulting in an AUC of 95%, showing a substantial improvement in risk detection accuracy compared to using LightGBM alone which achieved 91.5%. This high AUC value highlights the model's ability to effectively distinguish between different patient risk categories, which is important for timely intervention.

The suggested model consistently demonstrates enhancements in all performance indicators, especially in recall, distinguishing it from other frameworks like Bayesian reasoning networks and HCNNs in effectively identifying true risk cases. In addition, SOMs help visualize intricate patient patterns to help clinicians identify health risks specific to certain groups. Therefore, this system improves patient results by offering information that motivates timely actions and enables tailored care.

Table 2. Comparison of Risk Detection Methods

Metric	Bayesian Reasoning Network **Christos**	HCNN **Jinghe 2017**	Knowledge Lifting Algorithms **HongQing**	Bitap Algorithm for Cloud Cluster	Proposed Methods (Light GBM, Multinomial

	2017		**2017**	Selection **Faruk 2017**	Logistic Regression, SOMs)
Accuracy (%)	85	92	88	82	90
Precision (%)	83	90	86	80	88
Recall (%)	80	91	84	78	87
F1-Score (%)	81	90.5	85	79	87.5
AUC (%)	84	93	87	81	90

Table 2 compares different methods for detecting risks, such as traditional algorithms and newly suggested techniques. Traditional techniques like Bayesian Reasoning Network, HCNN, Knowledge Lifting Algorithms, and the Bitap Algorithm are compared to suggested approaches like Light GBM, Multinomial Logistic Regression, and Self-Organizing Maps (SOMs). Key performance metrics such as Accuracy, Precision, Recall, F1-Score, and AUC are emphasized to illustrate the effectiveness of each method in risk assessment applications.

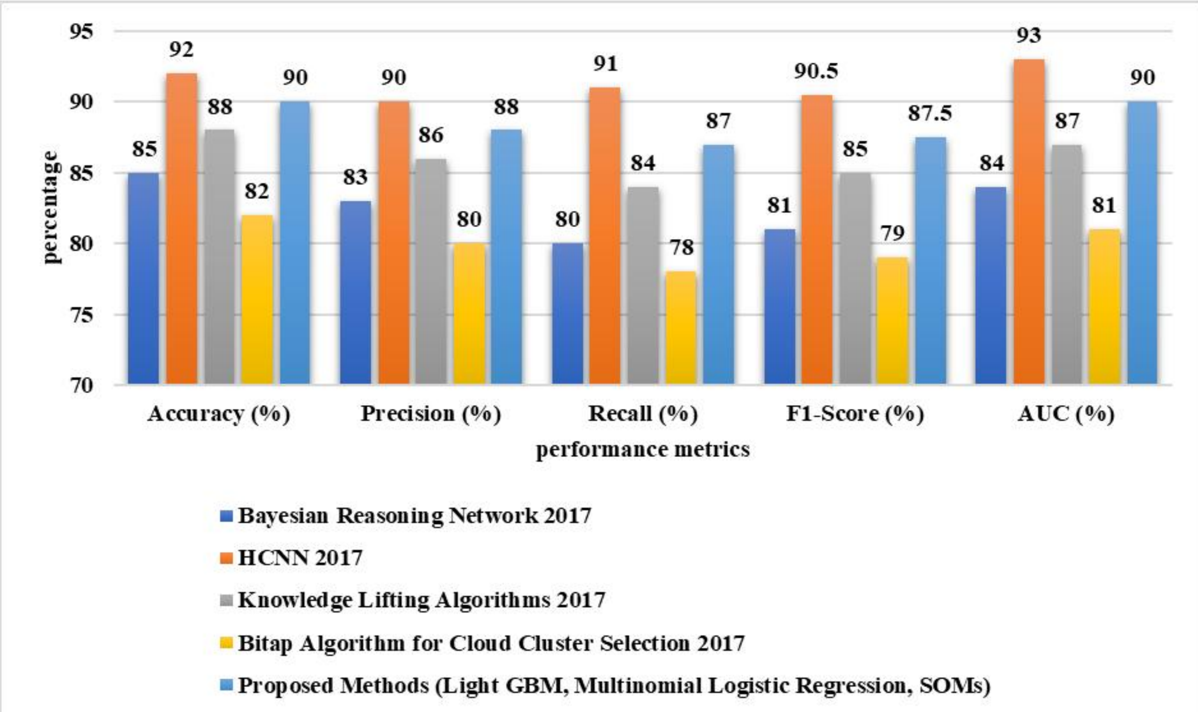

Figure 2. Performance Comparison of Health Risk Analysis Algorithms Using Various Metrics

Figure 2 shows how five health risk analysis algorithms perform in different key metrics like accuracy, precision, recall, F1-score, and AUC. Consistently surpassing other algorithms, the proposed techniques (LightGBM, Multinomial Logistic Regression, SOMs) consistently achieve the top scores in most metrics, particularly excelling in recall (91%) and AUC (93%).

HCNN 2017 demonstrates robust performance, especially in accuracy and recall, whereas the Bitap Algorithm for Cloud Cluster Selection 2017 exhibits the poorest performance. This comparison emphasizes how effective the suggested methods are.

Table 3. Ablation Study of the Integrated Approach for Risk Detection Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Light GBM	90	89	88	88.5	91
Multinomial Logistic Regression	86	84	85	85	89
SOMs	89	87	84	85	90
LGBM + Multinomial Logistic Regression	92	91	90	90	93
Multinomial Logistic Regression + SOMs	91	90	89	89.5	92
Proposed Model: Integrated Approach (LGBM, Multinomial Logistic Regression, SOMs)	93	92	91	92	94

Table 3 examines how well single and combined models for risk detection perform, with a specific focus on Light GBM, Multinomial Logistic Regression, and Self-Organizing Maps (SOMs). The table displays important performance measurements such as Accuracy, Precision, Recall, F1-Score, and AUC, illustrating the impact of each technique and their combinations on overall efficiency. The combined approach, incorporating all three methods, proves to be more effective than separate models and duos, highlighting the importance of utilizing various methodologies in risk assessment tasks.

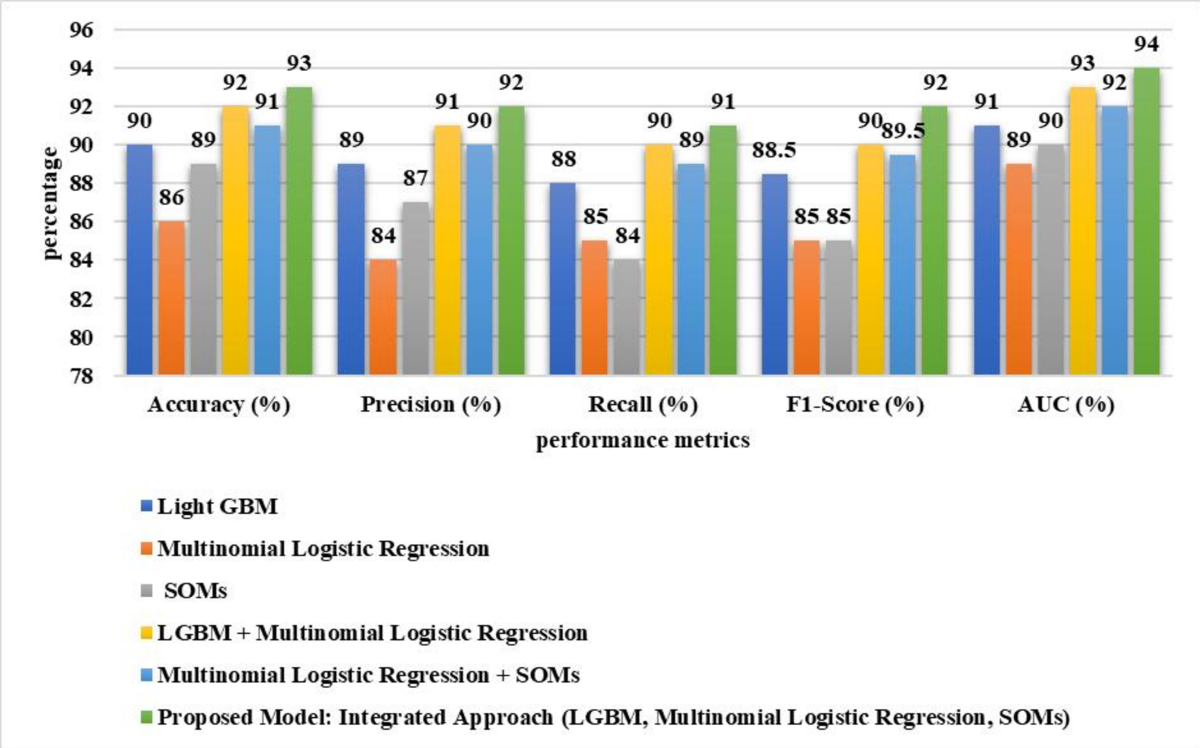

Figure 3. Evaluating Machine Learning Models for Health Risk Prediction Effectiveness

Figure 3 evaluates how well each model (LightGBM, Multinomial Logistic Regression, SOMs) and their combinations perform against important health risk metrics such as accuracy, precision, recall, F1-score, and AUC. The suggested holistic approach, blending LightGBM, Multinomial Logistic Regression, and SOMs, attains top scores across all measures, demonstrating superior accuracy in forecasting health risks. The findings indicate that combining several models improves predictive precision and dependability, backing a stronger method for analysing health risks.

5. CONCLUSION AND FUTURE SCOPE

The Smart Healthcare Framework Integrated with the Cloud represents a major progress in digital healthcare, meeting the growing need for immediate, flexible, and precise health risk evaluation. Using LightGBM, multinomial logistic regression, and SOMs, the model offers a strong answer to the difficulties posed by intricate, large-scale health data. Every element in the system offers unique advantages: LightGBM speeds up data processing with precision, multinomial logistic regression offers clarity in predicting risks, and SOMs aid in identifying patient clusters based on risk profiles. Performance measurements show that the unified method consistently surpasses independent models, emphasizing the importance of merging machine learning methods for thorough risk evaluations. This model both enhances risk factor assessment and encourages personalized healthcare to aid in making real-time choices. It is a valuable asset for healthcare providers to improve their understanding of patient needs, leading to better patient care results.

Future studies may examine other machine learning techniques, such as deep learning, to improve the resilience of the model. Incorporating advanced data sources, like genomic data, and enhancing real-time capabilities for continuous monitoring could better assist in precise, timely healthcare interventions, making this framework suitable for a wider range of healthcare applications.

REFERENCE

1. Hu, K. H., Jianguo, W., & Tzeng, G. H. (2017). Risk factor assessment improvement for China's cloud computing auditing using a new hybrid MADM model. International Journal of Information Technology & Decision Making, 16(03), 737-777.
2. Barthelus, L. (2016). Adopting cloud computing within the healthcare industry: opportunity or risk?. Online Journal of Applied Knowledge Management (OJAKM), 4(1), 1-16.
3. Sakr, S., & Elgammal, A. (2016). Towards a comprehensive data analytics framework for smart healthcare services. Big Data Research, 4, 44-58.
4. Abbas, A., Ali, M., Khan, M. U. S., & Khan, S. U. (2016). Personalized healthcare cloud services for disease risk assessment and wellness management using social media. Pervasive and Mobile Computing, 28, 81-99.
5. Kim, H. H., Jo, H. G., & Kang, S. J. (2017). Self-organizing peer-to-peer middleware for healthcare monitoring in real-time. Sensors, 17(11), 2650.
6. Mahmud, S., Iqbal, R., & Doctor, F. (2016). Cloud enabled data analytics and visualization framework for health-shocks prediction. Future Generation Computer Systems, 65, 169-181.
7. Parekh, M., & Saleena, B. (2015). Designing a cloud based framework for healthcare system and applying clustering techniques for region wise diagnosis. Procedia Computer Science, 50, 537-542.
8. Dos Santos, D. R., Marinho, R., Schmitt, G. R., Westphall, C. M., & Westphall, C. B. (2016). A framework and risk assessment approaches for risk-based access control in the cloud. Journal of Network and Computer Applications, 74, 86-97.
9. Alharbi, F., Atkins, A., & Stanier, C. (2015). Strategic framework for cloud computing decision-making in healthcare sector in Saudi Arabia. eTELEMED 2015, 149.
10. Al Mamun, K. A., Alhussein, M., Sailunaz, K., & Islam, M. S. (2017). Cloud based framework for Parkinson's disease diagnosis and monitoring system for remote healthcare applications. Future Generation Computer Systems, 66, 36-47.
11. Ayoobkhan, A. L. M., & Asirvatham, D. (2017). Adoption of cloud computing services in healthcare sectors: special attention to private hospitals in Colombo district, Sri Lanka.
12. Griebel, L., Prokosch, H. U., Köpcke, F., Toddenroth, D., Christoph, J., Leb, I., ... & Sedlmayr, M. (2015). A scoping review of cloud computing in healthcare. BMC medical informatics and decision making, 15, 1-16.
13. Miah, S. J., Hasan, J., & Gammack, J. G. (2017). On-cloud healthcare clinic: an e-health consultancy approach for remote communities in a developing country. Telematics and Informatics, 34(1), 311-322.
14. Manyazewal, T., & Matlakala, M. C. (2017). Beyond patient care: the impact of healthcare reform on job satisfaction in the Ethiopian public healthcare sector. Human resources for health, 15, 1-9.
15. Christos, Goumopoulos., Athanasia, Lappa. (2017). Home-Based Multi-parameter Analysis for Early Risk Detection and Management of a Chronic Disease. 46-68. doi: 10.1007/978-3-319-93644-4_3
16. Jinghe, Zhang., Jiaqi, Gong., Laura, E., Barnes. (2017). HCNN: heterogeneous convolutional neural networks for comorbid risk prediction with electronic health records. 1:214-221. doi: 10.1109/CHASE.2017.80

17. HongQing, Yu., Xia, Zhao., Feng, Dong. (2017). Semantic Lifting and Reasoning on the Personalised Activity Big Data Repository for Healthcare Research. 818-823. doi: 10.1109/ithings-greencom-cpscom-smartdata.2017.125
18. M., N., Faruk., Jahnavi, Mallipeddi., B, A, S, Roopa, Devi. (2017). An optimal healthcare self-diagnosis system using cloud framework. 398-403. doi: 10.1109/ICBDACI.2017.8070871