SehatSathi: A Hybrid Edge–AI Healthcare Assistant for Offline-First Community-Level Medical Support

Firdous Chandio, Muhammad Saleem Vighio, Bushra Khan
Department of Computer Science, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah
*Correspondence: chandiofirdous3@gmail.com

Citation | Chandio. F, Vighio. M. S, Khan. B, "SehatSathi: A Hybrid Edge–AI Healthcare Assistant for Offline-First Community-Level Medical Support", IJIST, Vol. 7 Issue. 10 pp 95-109, December 2025
DOI | https://doi.org/10.33411/IJIST/1727
Received | November 04, 2025 **Revised |** November 26, 2025 **Accepted |** November 29, 2025 **Published |** December 02 2025.

More than 60 percent of the population of Pakistan lives in rural or under-resourced areas, where timely access to effective healthcare is a significant challenge. This paper presents SehatSathi, a hybrid edge-AI integrated healthcare assistant designed with an offline-first architecture, supporting basic symptom triage, healthcare advice, and facilitating local doctor-patient appointment requests. The proposed system conducts symptom processing and disease prediction locally using multiple machine learning models for continuous operations in a low-connectivity environment. All the user interactions, including appointment requests, are persisted locally on the device and automatically synchronized when network access becomes available. The system supports Urdu, Sindhi, and English. A conversational chatbot powered by LLaMA can also be integrated through an API. Component-based modularity ensures that the system architecture preserves privacy, explainability, and graceful degradation when network connectivity is unavailable. The empirical assessment focuses on offline predictive performance, system responsiveness, and operational availability under low-connectivity conditions. Results show stable predictive performance and reliable usability even without internet connectivity.

Keywords: Edge AI, Offline Healthcare, Disease Prediction, Machine Learning, Multilingual Chatbot, Offline Symptom Triaging

Introduction:

Poor access to trusted health services is a major challenge in many developing countries, while access is further limited by poor internet connectivity, linguistic diversity, and resource constraints of healthcare providers in underserved and rural settings. In addition, poor language facilities seriously impede patients' communication with healthcare providers, despite Pakistan having a huge proportion of its population living in remote areas with poor access to health infrastructure facilities. These challenges require affordable and accessible digital healthcare services capable of operating in poorly structured environments. Recent breakthroughs in mobile health and artificial intelligence have paved the way for the development of advanced digital healthcare assistants, which can address most of the day-to-day medical queries and offer preliminary health advice. However, various existing systems depend largely on constant internet connectivity and cloud-based processing, thereby restricting their usability in low-connectivity environments. Moreover, most such systems offer support for English alone, which limits their efficiency in conditions that hold linguistic diversity. These are the drawbacks that impede the process of AI-based healthcare tool adoption in rural areas, which are in dire need of offline capability and multilingual support.

This paper proposes SehatSathi, a hybrid edge-AI healthcare assistant developed with an offline-first architecture to facilitate preliminary symptom assessment, healthcare guidance, and management of appointment requests. The proposed system locally performs symptom processing and disease prediction, using machine learning models, thereby allowing seamless operation without dependency on network connectivity. User interactions, including appointment requests, are cached locally on the device. They are automatically synchronized when internet connectivity is available, ensuring continuity in disconnected environments.

For better accessibility and user interaction, SehatSathi has a facility to communicate in Urdu, Sindhi, and English languages and incorporates a text chatbot based on LLaMA through application programming interface (API) services as a web option. By this, there will be a modular facility to provide a natural way of interaction through language, along with knowledge about health-related issues, and at the same time, the basic functionality will be intact, even in a network failure situation, ensuring privacy, explainability, and safe handling of health information, without providing clinical diagnosis, which will lead to early self-assessment and decision-making.

The major contributions made by this research can be outlined below

Design of a hybrid edge-AI healthcare assistant that prioritizes offline functionality to support communities with limited or unreliable internet connectivity.

Integration of a machine-learning–based symptom evaluation module with local data storage and deferred synchronization to enable reliable operation in low-connectivity environments.

Incorporation of a multilingual text-based chatbot powered by LLaMA, providing an interactive interface for user queries and healthcare guidance.

Comprehensive experimental evaluation of multiple machine learning models to assess system performance and reliability under limited connectivity scenarios.

Literature Review:

Artificial intelligence-powered healthcare assistants are known for making medical information and early health advice more accessible, especially through mobile health platforms. Early healthcare chatbots were mostly rule-based and followed set conversation paths, so they were less adaptable and did not understand context well [1]. Early healthcare chatbots, while offering basic health information, were limited in scaling and flexibility, reducing their effectiveness in real-world healthcare applications.

To address these issues, machine learning-based solutions are proposed in the context of symptom analysis and preliminary forecasting of diseases. Some research shows the benefits

of using supervised learning algorithms for early health diagnosis, analyzing symptoms in a structured format [2][3]. The systems are promising to regard predictability, but they require the use of cloud services and hence suffer from issues of inertia, privacy, and connectivity [4].

Recently, large language models have also been introduced into chatbots designed for the healthcare sector. Research has shown that large language models can produce more natural and contextually relevant answers compared to the traditional chatbot design [5][6]. However, several research studies have shown that the uncontrolled usage of large language models can result in several risks associated with hallucinations, bias, explainability, and the distribution of incorrect information and unethical behavior [7][8]. Therefore, it can be inferred from the literature that large language models must not be used in a diagnostic manner.

Language support for multiple languages is another area faced by the AI assistants for healthcare. Experimental studies reveal that most of the LLMs, as well as the chatbots for healthcare, perform better in English compared to other languages with limited resources [9]. Although the importance of the same is understood clearly, the use of languages such as Urdu or Sindhi is an unexplored area for the existing chatbot services for the healthcare domain [10].

Some of the most promising design principles include edge-centric architecture and offline-first methodologies. These are considered some of the efficient ways to reduce challenges that arise due to the demand for constant Internet connectivity. Previous studies show that machine learning based on edge can provide faster system responses and improve data privacy of information collected, especially when the network is unavailable [11].

Most healthcare chatbots cannot perform offline machine learning inference and rely on cloud-based processing for conversation. In summary, recent state-of-the-art studies have brought forth the potential of AI-based health care chatbots while also recognizing serious deficiencies with respect to dependence on connectivity, multilingual support, and safe integration of large language models. Different from these previous works, the SehatSathi proposed herein embodies a hybrid approach toward edge AI with application to health care in offline settings and provides a multilingual interface.

Summary of Contributions:

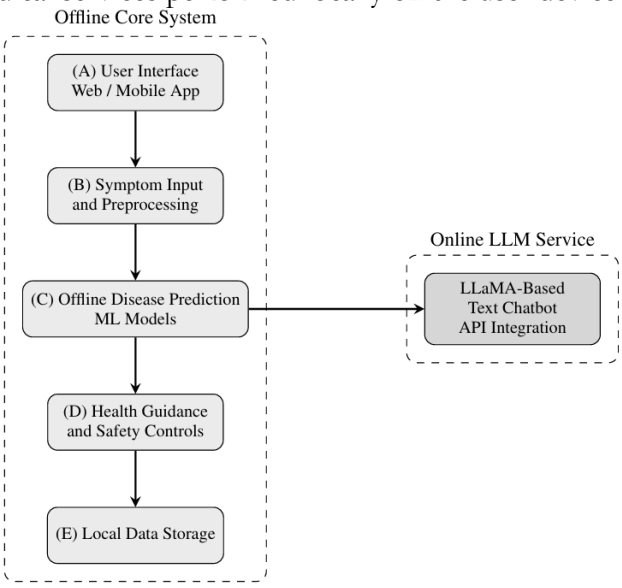
Table Key contributions of the proposed SehatSathi framework.

Table 1. Key contributions of the proposed SehatSathi framework.

Contribution	Description
Offline-first healthcare assistant	Enables symptom triage and health guidance without continuous internet connectivity.
Edge-based disease prediction	Local deployment of ML models for on-device disease prediction.
Comparative ML evaluation	Performance comparison of seven classifiers for symptom-based prediction.
Multilingual accessibility	Supports Urdu, Sindhi, and English for wider accessibility.
Hybrid AI interaction	An optional LLaMA-based chatbot for conversational guidance when online.
Local data persistence	Stores user data locally and synchronizes when the internet is available.
Graceful degradation	Maintains core features offline while disabling cloud services temporarily.

System Architecture Overview: The SehatSathi system is proposed with a hybrid edge AI architecture to ensure the availability of trust worthy medical assistance even under situations

with or without internet connectivity. The methodology prioritizes an offline-first design, with all core medical services performed locally on the user device.



As illustrated in Figure 1, the system architecture is divided into two major layers:

Offline Core System (Components A–E) is responsible for user interaction, symptom processing, disease prediction, health guidance, and local data storage.

Online LLM Service, consisting of a LLaMA-based text chatbot integrated via an API for conversational support.

Each methodological subsection corresponds directly to the labeled components (A–E) in the system architecture diagram.

Methodology:

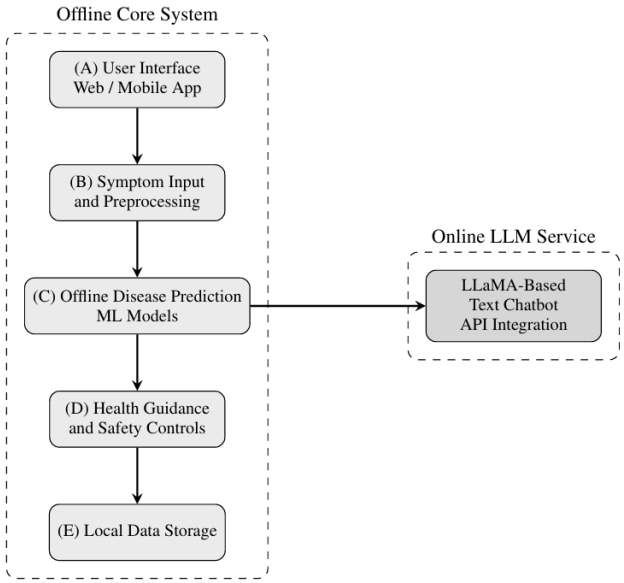


Figure 1. System architecture of the proposed SehatSathi

User Interface and Interaction Layer (Component A):

The user interface is developed as a web and mobile application and functions as the central hub for interaction between users and the system. The user interface provides a structured symptom selection and follows usability principles suitable for rural users and those with low literacy. Free-text clinical entry is excluded to simplify interactions.

The interface supports:

symptom selection, appointment booking requests, and access to conversational assistance All interactions are handled locally when internet connectivity is unavailable, ensuring uninterrupted access to core system functionalities.

Symptom Input Representation and Preprocessing (Component B):

User-reported symptoms are encoded into a structured, machine-readable format using binary feature representation.

Each symptom is represented as:

$$x_i = \begin{cases} 1, \text{if symptom } i \text{ is present} \\ 0, \text{otherwise} \end{cases}$$

This results in a fixed-length symptom vector

$$x = [x_1, x_2, \dots, x_n]$$

Binary encoding is used because it efficiently represents symptom data and is compatible with classical machine learning classifiers. Preprocessing steps include:

validation of symptom inputs,

alignment with the trained feature space,

handling missing or incomplete symptom selections.

The processed feature vector is then forwarded to the offline disease prediction engine.

Offline Disease Prediction Engine (Component C):

The disease prediction module can conduct local inference independently without using internet connections. To enhance robustness and reduce model-specific biases, the system employs seven supervised machine learning classifiers, each trained independently on identical symptom-disease data.

$$Let\ x \in \mathbb{R}^n\ denote\ the\ symptom\ vector\ and\ y \in \{1,2,\dots,K\} represent\ disease\ classes.$$

$$Each\ classifier\ learns\ a\ mapping: f(x) \to y$$

The following subsections describe each model in detail

Gaussian Naive Bayes Classifier:

Purpose: Probabilistic disease inference

Model Type: Generative classifier

Naive Bayes applies Bayes' theorem under the assumption that symptoms are conditionally independent given the disease class:

$$P(y \mid x) \propto P(x \mid y)P(y)$$

For Gaussian Naive Bayes, each feature likelihood is modeled as:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_{y,i}^2}} \exp\left(-\frac{(x_i - \mu_{y,i})^2}{2\sigma_{y,i}^2}\right)$$

The predicted disease corresponds to the class with the maximum posterior probability

Logistic Regression:

Purpose: Linear probabilistic classification

Model Type: Discriminative classifier

Logistic Regression models the conditional probability of disease classes using a linear combination of input features:

$$P(y = k \mid x) = \frac{e^{\beta_k^T x}}{\sum_{j=1}^{K} e^{\beta_j^T x}}$$

The Logistic Regression model is optimized by maximizing the log-likelihood function using gradient-based methods, providing stable and interpretable probability estimates.

Decision Tree Classifier:

Purpose: Rule-based clinical reasoning
Model Type: Non-parametric classifier
Decision Trees recursively partition the feature space by selecting symptom features that maximize information gain:

$$IG = H(S) - \sum_i \frac{|S_i|}{|S|} H(S_i)$$

Here, H(S)H(S)H(S) denotes entropy. Each leaf node corresponds to a disease prediction, providing transparent decision logic aligned with clinical reasoning

Random Forest Classifier:
Purpose: Ensemble robustness
Model Type: Bagging Ensemble

The Random Forest constructs an ensemble of TTT decision trees trained on bootstrapped samples, with the final prediction determined by majority voting:

$$\hat{y} = \text{mode}\{f_1(x), f_2(x), \ldots, f_T(x)\}$$

This approach reduces variance and improves generalization performance.
Gradient Boosting Classifier:
Purpose: Error-minimizing ensemble learning
Model Type: Boosting Ensemble

Gradient Boosting builds models sequentially, where each model minimizes the loss function of the previous ensemble:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$
$$where\ h_m is\ the\ weak\ learner\ and\ \gamma_m is\ the\ learning\ rate.$$

This model effectively captures complex symptom-disease relationships.
Support Vector Machine (SVM):
Purpose: Maximum-margin classification
Model Type: Discriminative classifier
SVM identifies a hyperplane that maximizes the margin between disease classes:

$$\min_{w,b} \quad \frac{1}{2}|w|^2 \text{ s.t. } y_i(w^\top x_i + b) \geq 1$$

This formulation improves generalization, especially in high-dimensional feature spaces.
Multilayer Perceptron (Neural Network):
Purpose: Non-linear feature learning
Model Type: Feed-forward neural network
The MLP consists of multiple hidden layers and computes:

$$h^{(l)} = \sigma\left(W^{(l)} h^{(l-1)} + b^{(l)}\right)$$

The model is trained using backpropagation to minimize cross-entropy loss, enabling it to capture higher-order symptom interactions.
Health Guidance and Safety Controls (Component D):

The Health Guidance and Safety Controls component acts as an intermediary layer between the offline disease prediction engine and the user interface, ensuring that prediction outputs are converted into safe, non-diagnostic, and appropriate health advice. This component generates initial health advice based on the predicted disease categories and the corresponding confidence levels, while avoiding unsafe predictions and overreliance on them. The structured health advice messages are generated using pre-defined templates that include a brief, non-technical description of the predicted disease, general advice on precautions such as monitoring symptoms, staying hydrated, or resting, and a recommendation to visit a qualified healthcare professional.

For the system to be ethical and safe for use, it will enforce strong output constraints. The system will refrain from generating any form of clinical decision-making output, including medication prescriptions, treatment plans, dosage recommendations, or any other form of prescription. Each guidance message includes a disclaimer that the output is for preliminary awareness only and does not constitute professional medical diagnosis or treatment. When symptom patterns or predicted disease categories show signs of potentially severe or high-risk diseases, the system will enforce rule-based safety measures that limit the scope of information and immediately advise users to seek medical attention. All the guidance generation and safety control measures will be entirely offline, relying on locally available predictions and predefined rules, to ensure the system operates without interruptions in low-connectivity environments while maintaining user privacy.

Local Data Storage and Offline-First Appointment Management (Component E):

To make the system continuously operational and user-friendly even in rural areas and places with low connectivity, SehatSathi uses a local storage mechanism for data as a primary design element. For storing all the user input data, such as symptoms, tentative predictions for diseases, and doctor and patient appointments, the system uses lightweight persistent storage. The appointment management subsystem follows an offline-first paradigm: offline requests are stored locally with doctor, time slot, and patient information, then synchronized when connectivity is restored.

As soon as the connectivity to the network is available, a synchronization process begins. The scheduling request saved on the local computer gets transmitted to the back-end server and performed in the order in which it was generated. The Deferred synchronization ensures accurate processing without requiring repeated user input. The user gets the impression of a seamless process for scheduling an appointment.

This approach directly addresses rural healthcare accessibility, ensuring continuity even when internet connectivity is unreliable.

LLaMA-Based Conversational Assistant (Online LLM Service):

SehatSathi also incorporates the LLaMA text chatbot to promote interactive functionality, especially among the linguistically diverse population, as an optional online component. The chatbot has an application programming interface (API) through which the functionality of the chatbot is accessible only if the internet connection is available.

The communication assistant has been developed solely for non-clinical purposes, which include:

explaining system outputs in natural language.

responding to general health-related queries.

providing multilingual support in Urdu, Sindhi, and English.

In fact, the LLaMA-based chatbot does not engage in the analysis of patient symptoms or predictions regarding patient outcomes. All matters related to inference in medicine are carried out solely through the offline machine learning algorithms, which are discussed in Section C.

Isolating the LLM component preserves reliability, safety, and explainability in prediction tasks, while still providing improved conversational usability when online.

Offline System Operation and Graceful Degradation:

SehatSathi employs a graceful degradation strategy to manage transitions between offline and online operational states. This scheme is always observing network connectivity for the dynamic enabling/disabling of any online facility.

When offline:

symptom checking,

disease prediction,

health guidance,

and appointment scheduling continues to function normally using local resources.
When online:
Conversational assistance via the LLaMA chatbot becomes available,
locally stored appointment data is synchronized,
System content is optionally updated.

This coordination strategy ensures that no core health care functionality depends on continuous internet access, making the system resilient against connectivity disruptions. From an architectural perspective, this aligns with edge computing principles. Computation and decision-making occur close to the user, reducing latency and improving privacy and availability.

Results and Performance Evaluation:

This section presents the experimental results of the proposed offline symptom-based disease prediction module. Multiple machine learning models were evaluated to assess predictive performance, robustness, and suitability for deployment in offline, resource-constrained healthcare environments. The evaluation focuses on standard classification metrics, including accuracy, precision, recall, and F1-score.

Experimental Setup:

Each model was trained and evaluated on a preprocessed symptom-disease dataset, with symptoms represented as binary feature vectors indicating presence or absence. The data were split into training, validation, and test sets to assess model performance without bias. Results were reported on the test set.
The following machine learning models were evaluated as part of the offline disease prediction component (Module C in Fig. 1):
Logistic Regression
Decision Tree
Gradient Boosting
Support Vector Machine (SVM)
Naive Bayes
Neural Network (Multi-Layer Perceptron)

These models were selected to represent a diverse range of linear, probabilistic, tree-based, ensemble, and nonlinear learning approaches.

Accuracy Comparison Across Models:

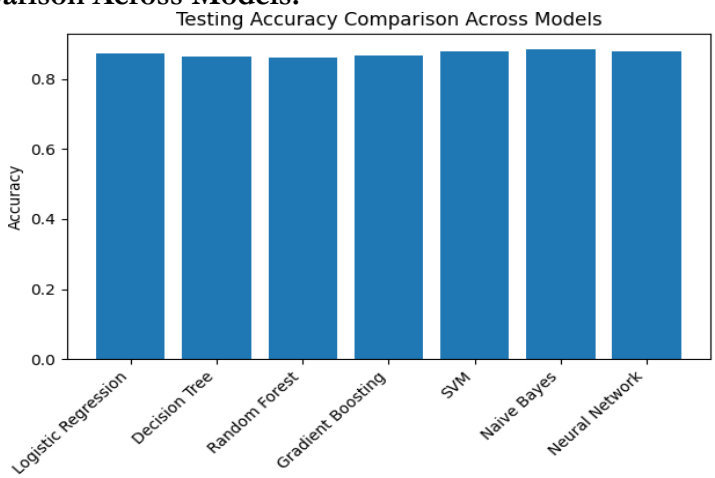


Figure 2. Comparison of testing accuracy across the evaluated machine learning models using the symptom-based healthcare dataset employed in the SehatSathi framework.

Figure 2 shows the test accuracy of each model. The results show that most models perform similarly, with accuracies ranging from approximately 85% to 89%.

Naive Bayes and SVM achieved the highest accuracies, demonstrating strong performance in symptom-based classification tasks. Other models—including Logistic Regression, Random Forest, Gradient Boosting, and Neural Network—also achieved relatively high accuracy, suggesting that symptom-driven disease prediction can be effectively modeled using multiple learning paradigms.

The Decision Tree showed slightly lower accuracy compared to the other models.

Key observation: The results indicate consistently high accuracy across the evaluated models.

Model Ranking Based on Test Accuracy:

Figure 3 shows the comparison of different models ranked against their test accuracy. Naive Bayes ranks highest, followed closely by SVM and Neural Network models. Ensemble methods, including Random Forest and Gradient Boosting, also achieved strong rankings, demonstrating their robustness.

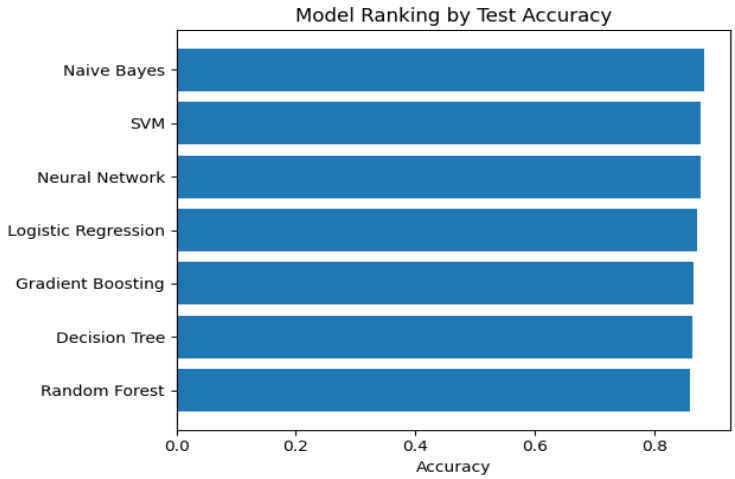


Figure 3. Comparative ranking of evaluated machine learning models based on their testing accuracy on the symptom-based healthcare dataset used in the SehatSathi framework.

This ranking points to one of the most important design insights: simple probabilistic models, such as Naive Bayes, can outperform more complex ones when the assumption of symptom independence approximately holds. As a result, they are also well-suited for offline and low-resource settings.

Precision, Recall, and F1-Score Analysis:

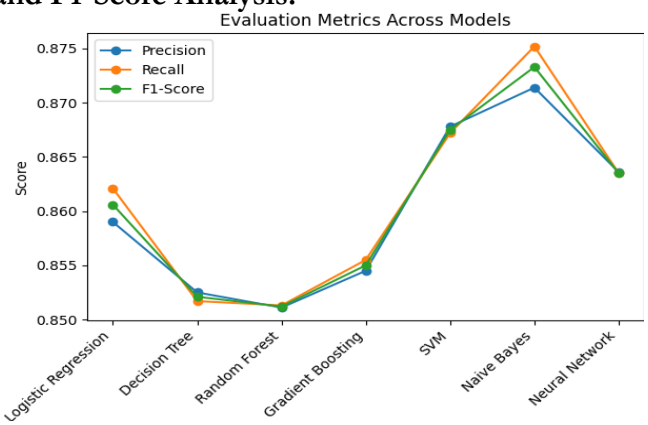


Figure 4. Evaluation metrics comparison across the evaluated machine learning models, including accuracy, precision, recall, and F1-score, using the symptom-based healthcare dataset in the SehatSathi framework.

Figure 4 illustrates the comparison of precision, recall, and F1 scores for each model tested, where the costs of false positives and false negatives may differ.

Precision reflects the correctness of predicted diseases.
Recall measures the ability to correctly identify actual disease cases.
F1-score provides a balanced measure between precision and recall.
The results indicate that Naive Bayes and SVM effectively balance all three metrics. Logistic Regression and Neural Network models maintain high F1 scores across the three metrics. Decision Tree and Random Forest algorithms show slightly lower recall and F1 scores, likely due to sensitivity to feature splitting and dataset size.

Comparative Performance Summary:

Table 2. Performance comparison of evaluated machine learning models based on accuracy, precision, recall, F1-score, and computational metrics on the symptom-based healthcare dataset used in the SehatSathi framework.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.8712	0.8590	0.8620	0.8605
Decision Tree	0.8647	0.8523	0.8515	0.8519
Random Forest	0.8606	0.8510	0.8510	0.8510
Gradient Boosting	0.8655	0.8545	0.8555	0.8550
Support Vector Machine (SVM)	0.8784	0.8678	0.8676	0.8677
Naive Bayes	**0.8836**	**0.8712**	**0.8750**	**0.8731**
Neural Network	0.8782	0.8635	0.8634	0.8634

Table 2 summarizes the overall performance of all models evaluated using accuracy, precision, recall, and F1-score. All the models attain an "Excellent" performance rating, which assures the viability of implementing the disease prediction module in an offline healthcare assistant.

The use of multiple models provides architectural flexibility, allowing the system to either:
Select the best-performing model for deployment, or
Use an ensemble or confidence-based agreement strategy to enhance reliability.

Discussion and System Implications:

These experimental results demonstrate that the proposed hybrid edge-AI framework can offer reliable disease prediction without continuous connectivity to the internet. Of special note is the strong performance of lightweight models, such as Naive Bayes and Logistic Regression, for offline deployment on mobile devices in rural and underserved areas.

This validates the system architecture presented in Fig. 1, where offline disease prediction works independently and the LLaMA-based chatbot enhances user interaction only if connectivity is available. This configuration provides graceful degradation and ensures privacy with continuous system availability

Key Takeaways:

Offline symptom-based disease prediction is accurate and reliable
Simple models perform competitively with complex ones
The system is suitable for edge deployment
Results validate the hybrid offline–online architecture

Discussion:

The results for the offline disease-prediction module prove that it is feasible to deploy a competent symptom-based healthcare assistant in an environment with limited connectivity and resources. For all the machine learning models tested, the system puts in strong performance, which means the symptom representation and preprocessing workflow used captures the dominant disease patterns. Naive Bayes and Support Vector Machine emerge as the top models for accuracy, precision, recall, and F1 score. These excellent performances of Naive Bayes are owed to its probabilistic framework and suitability for high-dimensional, sparse symptom data, whereby each symptom independently contributes toward the likelihood

of a disease. This characteristic property makes Naive Bayes particularly suitable for offline healthcare applications, as it brings together high accuracy for low computational overhead and fast inference time compatible with system edge-AI design objectives.

SVM also performs competitively, benefiting from its ability to create strong decision boundaries in high-dimensional feature spaces. The strong generalization capability allows it to effectively separate overlapping symptom patterns. However, SVM requires more computational resources during training and model tuning, which may limit scalability for periodic offline retraining on edge devices.

Other ensemble models, namely Random Forest and Gradient Boosting, also perform stably in the prediction task. These ensemble models utilize multiple decision trees to reduce variance and capture complex interactions among symptoms. While they reach a high accuracy, their increased memory requirements and inference complexity make them less appropriate for fully offline on-device deployment in low-end devices compared to the simpler probabilistic models.

The Decision Tree model proved marginally worse than ensemble and probabilistic methods. This is somewhat expected, as single-tree methods are often prone to over-fitting and may fail to generalize to diverse symptom combinations. However, Decision Trees do offer high interpretability, which is an asset in health applications that rely on explainability.

The Neural Network model also achieved strong accuracy, emphasizing its ability to grasp the relationship between symptoms and disease outcomes. However, these types of models tend to require intensive tuning, larger datasets, and heavier processing power. Thus, the application of these models in an offline setting may necessitate resorting to optimizations such as model compression, which should serve as directions for future aspects of this study.

Another key observation is that differences among model performances are relatively small, which means that predicting disease from symptoms for common diseases is a well-defined classification problem. It is consistent with why the system chooses to put emphasis on offline reliability over cloud inference capability. Overall, the system is adaptable in selecting models based on capability, maintaining a balance between efficiency and accuracy.

An optional online LLaMA-based chatbot complements the offline prediction module without compromising system availability. It improves the user interface, explanation, and support for multiple languages when connectivity is available, whereas its core diagnosis and management features are available even without internet connectivity.

In general, the discussion raises the point that the hybrid edge-AI healthcare assistant provides a good blend of predictive accuracy, interpretability, and availability. The results favor the appropriateness of the system for underserved and rural areas where dependable offline health care support is a priority. Future work will focus on increasing disease coverage, enhancing multilingual natural language interaction, and real-world clinical validation.

Conclusion:

This work presents SehatSathi, a hybrid edge–AI health assistant that aims at minimizing these accessibility barriers in resource-constrained, low-connectivity areas. This system runs on an offline-first architecture to support seamless symptom assessment, preliminary disease prediction, and management of doctor–patient appointments with no need for continual access to the internet. The proposed framework strikes a balance between reliability, accessibility, and usability by integrating lightweight machine learning models for offline inference with an optional LLaMA-based conversational component to enhance user interaction.

Evaluation of the offline disease prediction module shows that a variety of classical and ensemble machine learning models produce consistent and dependable performance on structured symptom-based datasets. The experimental evaluation demonstrated strong performance across models, with Naive Bayes achieving the highest accuracy (0.8836),

followed by Support Vector Machines (0.8784) and Neural Networks (0.8782). The findings presented show that probabilistic and margin-based classifiers are particularly suitable for deployment in resource-constrained environments due to their balance between accuracy and computational efficiency.

The findings presented show that the probabilistic and margin-based classifiers, especially Naive Bayes and Support Vector Machines, are very suitable for deployments in resource-constrained environments due to their good balance between accuracy and computational efficiency. System-level verification further confirms the robustness of the architecture by showing seamless offline operations and appointment data synchronization once network connectivity is restored.

Taken together, these findings demonstrate that a hybrid edge–AI health assistant can be deployed to support early self-assessment, increase access to healthcare, and enhance digital health in linguistically and infrastructurally diverse environments.

Limitations:

The proposed system has several limitations despite the promising results: First, the disease prediction module is evaluated on a structured symptom–disease dataset, which may not fully capture the variability, ambiguity, and noise characteristic of real-world patient-reported symptoms. Thus, the reported results reflect dataset-level performance rather than clinical diagnostic accuracy.

Second, the system focuses on preliminary disease prediction only for common conditions and does not replace professional medical diagnosis or treatment. The lack of real-world clinical validation and user studies limits the evaluation of long-term efficacy, user trust, and clinical safety.

Third, while the LLaMA-based chatbot improves user interaction, it is an online-only component and is not available for use offline. The chatbot also intentionally does not take part in clinical decision-making due to safety concerns and may limit its ability to carry out advanced medical reasoning.

Finally, the current implementation assumes ample device storage and computational capacity for offline model execution, which may not generalize across hardware platforms.

Future Work:

Future research will address identified limitations and extend the capabilities of the proposed system. Clinical validation, in collaboration with healthcare professionals and through controlled user studies, will be prioritized to assess real-world effectiveness and usability. The expansion of the symptom–disease dataset for a wide spectrum of conditions and more diversified patient profiles should further improve the generalization of the model.

Model optimization techniques, such as pruning, quantization, and knowledge distillation, will be pursued in the future to enable more efficiency on low-resource devices. Further confidence estimation and uncertainty-aware prediction might increase decision reliability in cases of ambiguity.

Further development of multilingual and voice-based interaction will add offline speech-to-text and text-to-speech capabilities, improving accessibility for low-literacy users. Lastly, deeper integration of telemedicine services and secure data synchronization mechanisms is foreseen to further reinforce the system's role in community-level healthcare support.

Appendix A: Dataset Description and Symptom Encoding:

The disease prediction component uses a structured symptom-disease dataset that includes a wide variety of common medical conditions together with their respective symptom profiles. Each patient instance is represented as a binary feature vector where each feature indicates whether the corresponding symptom is present or absent. This allows for uniform input from all machine learning models, and this can be computed efficiently offline.

Let $x = [x1, x2, \ldots, xn]$

denote the symptom vector, where:

$$x_i = \begin{cases} 1, \text{if symptom } i \text{ is present} \\ 0, \text{otherwise} \end{cases}$$

The target label is a discrete disease class, hence treating the problem as multiclass classification. The dataset has been cleaned to remove inconsistent entries and then normalized to make it compatible for training probabilistic, distance-based, and ensemble learning algorithms.

Appendix B: Machine Learning Models and Hyperparameter Configuration:

All the models are trained and tested using the same dataset and the same feature representation to perform a comparative analysis with fairness. The offline disease prediction module utilizes the following models and their hyperparameters:

Naive Bayes (Gaussian):
Prior probability estimation enabled
Assumes conditional independence among symptoms
No manual hyperparameter tuning required

Support Vector Machine (SVM):
Kernel: Radial Basis Function (RBF
Regularization parameter C: tuned using cross-validation
Kernel coefficient γ: automatically scaled

k-Nearest Neighbors (kNN):
Number of neighbors k: empirically selected
Distance metric: Euclidean distance
Uniform weighting of neighbors

Decision Tree:
Splitting criterion: Gini impurity
Maximum depth: constrained to reduce overfitting
Minimum samples per leaf node enforced

Random Forest:
Number of trees: optimized for accuracy, efficiency tradeoff
Bootstrapped sampling enabled
Feature subset selection at each split

Gradient Boosting:
Learning rate: a small value for stable convergence
Number of estimators: empirically determined
Tree depth limited to prevent overfitting

Neural Network (Multi-Layer Perceptron):
Hidden layers: fully connected
Activation function: ReLU
Optimizer: Adam
Early stopping is used to prevent overfitting
All models are trained offline and serialized for deployment within the edge environment.

Appendix C: Offline System Operation and Data Synchronization:

The SehatSathi system follows an offline-first operational paradigm. Core functionalities, including symptom entry, disease prediction, and doctor–patient appointment scheduling, function exclusively without network connectivity. All user interactions and generated records are stored locally on the device via a lightweight embedded database.

A background synchronization will start when network connectivity is restored, securely pushing the locally stored appointment information to the central server. The result is a delayed but reliable appointment confirmation while keeping offline usability intact.

This system is designed to degrade gracefully in conditions of limited connectivity, ensuring continuous health care support regardless of network availability.

Appendix D: Evaluation Metrics:

To assess model performance, the following standard classification metrics are used:
Accuracy: proportion of correctly classified instances
Precision: reliability of positive predictions
Recall: ability to identify relevant disease cases
F1-score: harmonic mean of precision and recall

These metrics provide a balanced evaluation framework, particularly important for healthcare applications where false positives and false negatives have different implications.

Appendix E: Ethical Considerations and Safety Disclaimer:

It is designed to provide a preliminary health-oriented diction and should not be regarded either as a professional diagnosis or medical treatment. And all the predictions made would carry clear disclaimers, prompting users to seek verification by appropriate follow-up with qualified healthcare professionals.

The LLaMA-powered chatbot module does not allow clinical decisions and can only provide conversational support for information clarification. No personally identifiable health data is transmitted without the user's consent, thereby preserving privacy and ensuring adherence to the norms of ethics.

Conflict of Interest:

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References:

[1] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinagaran, "Conversational Agents in Health Care: Scoping Review and Conceptual Analysis," *JMIR Publ.*, vol. 22, no. 8, 2020, [Online]. Available: https://www.jmir.org/2020/8/e17158/

[2] Vishal Pal, Parvej Aalam, "Symptom-Based Disease Prediction Using Machine Learning: A Web Application Approach," *Optim. Artif. Intell. Strateg. Eng. Manag.*, 2024, [Online]. Available: https://www.researchgate.net/publication/385162172_SymptomBased_Disease_Prediction_Using_Machine_Learning_A_Web_Application_Approach

[3] "Disease Prediction Using Machine Learning - GeeksforGeeks." Accessed: Feb. 09, 2026. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/disease-prediction-using-machine-learning/

[4] Mohammad Mehrtak, Seyedahmad Seyedalinaghi, "Security challenges and solutions using healthcare cloud computing," *J. Med. Life*, vol. 14, no. 4, pp. 448–461, 2021, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8485370/

[5] Christian Winkler, "Leveraging Large Language Models in Healthcare: From Speech Documentation to Conversational Agents," *Adv. Digit. Heal. Care*, pp. 187–206, 2026, [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-032-16837-5_16

[6] N. M. Bright Huo, Amy Boyle, "Large Language Models for Chatbot Health Advice Studies A Systematic Review," *JAMA Netw. Open*, 2025, [Online]. Available: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2829839

[7] Javad Pool, Marta Indulska, Shazia Sadiq, "Large language models and generative AI in telehealth: a responsible use lens," *J Am Med Inf. Assoc*, vol. 31, no. 9, pp. 2125–2136, 2024, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/38441296/

[8] James C. L. Chow, Kay Li, "Large Language Models in Medical Chatbots: Opportunities, Challenges, and the Need to Address AI Risks," *Information*, vol. 16, no. 7, p. 549, 2025, [Online]. Available: https://www.mdpi.com/2078-2489/16/7/549

[9] S. K. Yiqiao Jin, Mohit Chandra, Gaurav Verma, Yibo Hu, Munmun De Choudhury, "Better to Ask in English: Cross-Lingual Evaluation of Large Language Models for Healthcare Queries," *arXiv:2310.13132*, 2023, [Online]. Available: https://arxiv.org/abs/2310.13132

[10] D. N. H. K.-U.-R. K. Z. B. Nazish Basir*, "Leveraging Machine-Labeled Data and Cross-Lingual Transfer for NER in Urdu and Sindhi," *J. Inf. Commun. btn btn-dark btn-xs btn-round)*, vol. 19, no. 1.

[11] M. M. Kamruzzaman, Ibrahim Alrashdi, "New Opportunities, Challenges, and Applications of Edge-AI for Connected Healthcare in Internet of Medical Things for Smart Cities," *J. Healthc. Eng.*, 2022, [Online]. Available: https://www.researchgate.net/publication/358824613_New_Opportunities_Challenges_and_Applications_of_EdgeAI_for_Connected_Healthcare_in_Internet_of_Medical_Things_for_Smart_Cities



Copyright © by authors and 50Sea. This work is licensed under the Creative Commons Attribution 4.0 International License.