AI health assistant combining transformers and XGBoost for multilingual care

Shamiso Simango, Mainford Mutandavari
School of Information Science and Technology, Harare Institute of Technology, Harare, Zimbabwe

Article Info

Article history:

Received Jun 12, 2025
Revised Jun 16, 2026
Accepted Jul 13, 2026

Keywords:

Artificial intelligence
Low resource settings
Multilingual
No language left behind
Transformers
Virtual health assistant
XGBoost

ABSTRACT

Limited healthcare access, shortages of healthcare professionals, and linguistic diversity continue to impede timely symptom assessment and healthcare delivery in low-resource settings such as Zimbabwe. Existing virtual health assistant (VHAs) are frequently cloud-dependent, English-centric, and lack interpretable decision-making, limiting their effectiveness in bandwidth-constrained and privacy-sensitive environments. This study proposes CIMAS HealthMate, a hybrid multilingual VHA that integrates transformer-based natural language processing (NLP) with an explainable extreme gradient boosting (XGBoost) decision model to provide accurate and transparent symptom triage. The framework employs the no language left behind (NLLB) model for offline English–Shona translation, bidirectional encoder representations from transformers (BERT)-based models for intent classification and medical entity recognition, and XGBoost for structured triage recommendation. The system was evaluated using a multilingual symptom corpus and an anonymized electronic health record-style dataset comprising approximately 23,000 patient records. Experimental results achieved translation accuracies of 76.5% for Shona-to-English and 82.2% for English-to-Shona, symptom extraction accuracy of 86.6%, and end-to-end triage accuracy of 93.3% with an F1-score of 93.3%. These findings demonstrate that the proposed hybrid architecture effectively combines multilingual language understanding, interpretable machine learning, and offline deployment to deliver reliable and privacy-preserving triage support. The proposed approach provides a scalable and practical solution for improving equitable digital healthcare services in multilingual, resource-constrained environments.

Corresponding Author:

Shamiso Simango
School of Information Science and Technology, Harare Institute of Technology
Belvedere, Harare, Zimbabwe
Email: shamisosimaz@gmail.com

1. INTRODUCTION

Artificial intelligence (AI) has emerged as a transformative technology in digital healthcare, enabling conversational agents that can interpret patient-reported symptoms and provide timely clinical guidance. Recent advances in transformer-based architectures have significantly improved natural language understanding, multilingual communication, and human-computer interaction, making AI-powered virtual health assistants (VHAs) increasingly effective for healthcare support [1]. Explainable machine learning algorithms such as extreme gradient boosting (XGBoost) have demonstrated excellent predictive performance on structured clinical data while maintaining model interpretability [2]. Likewise, bidirectional encoder representations from transformers (BERT) have substantially enhanced contextual language understanding for healthcare-related natural language processing (NLP) tasks [3], [4]. Recent surveys further highlight the effectiveness of deep learning and prompt-based transformer models in medical text processing and healthcare applications [5]–[8].

Hybrid machine learning models that combine deep language understanding with structured prediction have demonstrated improved predictive performance in healthcare decision support [9], [10]. Furthermore, multilingual transformer-based medical chatbots have shown considerable potential for improving communication between patients and healthcare providers across different languages [11]. Advances in scalable transformer architectures and transfer learning have further expanded the capability of AI systems to process multilingual clinical information efficiently [12]–[14].

Despite these technological advances, healthcare delivery in many low-resource countries, including Zimbabwe, continues to face substantial challenges. Shortages of healthcare professionals, inadequate digital infrastructure, geographical barriers, and limited access to timely medical consultation continue to reduce healthcare accessibility. These challenges are further compounded by linguistic diversity, as many patients communicate primarily in indigenous languages such as Shona rather than English. Consequently, English-only digital health applications often fail to provide equitable healthcare services, particularly in rural and underserved communities. Previous studies have emphasized the importance of multilingual and low-resource digital health solutions for improving healthcare accessibility in developing regions [15], [16].

Recent developments in lightweight transformer models such as a lite BERT (ALBERT) and multilingual representation learning have improved language understanding while reducing computational complexity [17], [18]. Nevertheless, many existing VHAs remain cloud-dependent and operate as black-box systems, raising concerns regarding explainability, privacy, and ethical deployment in healthcare [19], [20]. Moreover, robust evaluation metrics and explainable AI are essential for ensuring reliable clinical decision support [21]–[23]. These limitations indicate the need for an intelligent healthcare assistant that combines multilingual language understanding, interpretable decision-making, and efficient offline deployment within a unified framework.

To address these challenges, this study proposes CIMAS HealthMate, a multilingual VHA designed for low-resource healthcare settings in Zimbabwe. The proposed framework integrates multilingual transformer-based NLP with an explainable XGBoost classifier to provide evidence-based symptom triage rather than medical diagnosis. The system incorporates offline English–Shona translation, BERT-based intent classification and medical entity recognition, and an interpretable decision layer to support transparent clinical recommendations. Furthermore, the proposed framework adopts a privacy-by-design architecture that aligns with international data protection principles and digital health recommendations [24], [25].

The main contributions of this study are fourfold. First, a hybrid multilingual VHA architecture is developed by integrating transformer-based NLP with an explainable XGBoost decision model. Second, the framework supports offline multilingual communication suitable for bandwidth-constrained environments. Third, the proposed architecture emphasizes interpretable and privacy-preserving AI for healthcare applications. Finally, the proposed system is evaluated using multilingual symptom data and structured electronic health record (EHR)-style data to demonstrate its effectiveness for accurate, transparent, and scalable multilingual symptom triage.

2. RELATED WORKS

Recent advances in AI have accelerated the development of VHAs for symptom assessment, patient engagement, and clinical decision support. Commercial platforms such as Babylon Health, Ada Health, and MobiHealth Kenya have demonstrated the potential of AI-assisted healthcare by providing conversational symptom assessment and health recommendations. However, most existing platforms rely on cloud-based infrastructures, require continuous internet connectivity, and primarily support widely spoken languages such as English. These characteristics limit their applicability in low-resource healthcare environments, where network connectivity is unreliable and many users communicate in indigenous languages. Furthermore, several commercial VHAs generate diagnostic suggestions, raising concerns regarding explainability, clinical accountability, and regulatory compliance.

From a methodological perspective, recent studies have shown that transformer-based language models, including BERT and its multilingual variants, substantially improve intent classification, medical entity recognition, and contextual language understanding through contextual embeddings. These models have become the foundation of modern multilingual conversational AI because they capture semantic relationships more effectively than conventional recurrent neural networks or rule-based approaches. Nevertheless, transformer models generally require substantial computational resources and often operate as black-box systems, limiting their deployment in privacy-sensitive and resource-constrained healthcare environments.

In parallel, ensemble learning algorithms such as XGBoost have consistently demonstrated excellent predictive performance for structured clinical data while providing interpretable feature importance and efficient inference. Consequently, XGBoost has been widely adopted for disease prediction, patient risk stratification, and clinical decision support. Unlike deep neural networks, gradient boosting models provide greater transparency, making them more suitable for evidence-based healthcare applications where explainability is essential.

Recent research has begun exploring hybrid architectures that combine transformer-based language understanding with structured machine learning models to leverage the complementary strengths of both approaches. While these hybrid systems improve conversational intelligence and prediction accuracy, relatively few studies have focused on multilingual healthcare applications for African languages or considered offline deployment, low-bandwidth operation, and compliance with regional data protection regulations. These limitations motivate the development of CIMAS HealthMate, which integrates multilingual transformer-based NLP with an explainable XGBoost decision model to provide privacy-preserving, multilingual symptom triage for low-resource healthcare settings. Table 1 compares the proposed CIMAS HealthMate framework with representative healthcare platforms. Unlike existing systems, the proposed framework combines multilingual support, offline operation, interpretable machine learning, and privacy-preserving deployment within a unified architecture, making it more suitable for multilingual, resource-constrained healthcare environments.

Table 1. Comparison of features between the CIMAS HealthMate hybrid VHA and existing health platforms (Babylon Health, Ada Health, MobiHealth Kenya)

Criterion	Proposed CIMAS HealthMate	Babylon Health	Ada Health	MobiHealth Kenya
NLP model	Transformer (BERT+no language left behind)	Proprietary neural AI	Proprietary AI	Rule-based/ machine learning
Decision model	Explainable XGBoost	Proprietary	Proprietary	Rule-based
Multilingual support	English–Shona	Limited	Multiple languages	English–Swahili
Offline capability	Yes	No	No	Partial
Explainable AI	Yes	Limited	Limited	Moderate
Symptom triage	Yes	Yes	Yes	Basic
Medical diagnosis	No	Yes	Yes	No
Privacy-by-design	Yes	Partial	Partial	Partial
Low-bandwidth deployment	Yes	No	No	Partial
Regulatory compliance	General data protection regulation, health insurance portability and accountability act, Zimbabwe cyber and data protection act	Health insurance portability and accountability act	General data protection regulation	Local regulations

3. METHOD

3.1. Overall system architecture

The proposed CIMAS HealthMate framework adopts a modular hybrid architecture consisting of three functional layers: translation, contextual understanding, and clinical decision-making, as illustrated in Figure 1. This layered design combines multilingual NLP with explainable machine learning to provide accurate, interpretable, and privacy-preserving symptom triage suitable for low-resource healthcare environments. The workflow begins with the translation layer, where user queries are automatically translated between Shona and English using the offline no language left behind (NLLB)-200 model. Offline translation eliminates dependence on cloud-based services, thereby improving system reliability, reducing communication latency, and preserving patient privacy in bandwidth-constrained environments.

The translated text is subsequently processed by the contextual understanding layer, which employs a BERT-based NLP pipeline. This layer performs intent classification to identify the user's healthcare request and medical entity recognition to extract clinically relevant information, including reported symptoms, symptom duration, severity, medications, and pre-existing medical conditions. The extracted entities are transformed into structured feature vectors suitable for downstream machine learning analysis. Finally, the structured clinical features are forwarded to the clinical decision-making layer, where an explainable XGBoost classifier predicts one of three evidence-based triage recommendations: home_remedy, wait_and_monitor, or see_doctor. Unlike diagnostic systems that attempt to identify specific diseases, the proposed framework focuses exclusively on symptom triage to support timely healthcare guidance while maintaining clinician oversight. The use of XGBoost also enables feature importance analysis, providing greater transparency and interpretability than end-to-end neural decision models.

The integration of multilingual translation, contextual language understanding, and interpretable machine learning enables CIMAS HealthMate to provide reliable, transparent, and clinically meaningful symptom triage while supporting offline deployment, multilingual communication, and privacy-preserving healthcare services in resource-constrained settings. By combining the contextual reasoning capabilities of transformer-based language models with the predictive accuracy and explainability of XGBoost, the proposed framework effectively bridges the gap between unstructured patient narratives and structured clinical decision support. Furthermore, its modular architecture facilitates scalability, allowing additional languages, clinical knowledge bases, and decision models to be incorporated with minimal system modification. The offline implementation reduces dependence on cloud infrastructure, ensuring consistent performance in low-bandwidth environments while safeguarding sensitive patient information. Consequently, CIMAS HealthMate offers a practical and extensible platform for improving equitable access to digital healthcare, reducing clinician workload through evidence-based triage, and supporting informed healthcare decision-making in multilingual, low-resource healthcare environments.

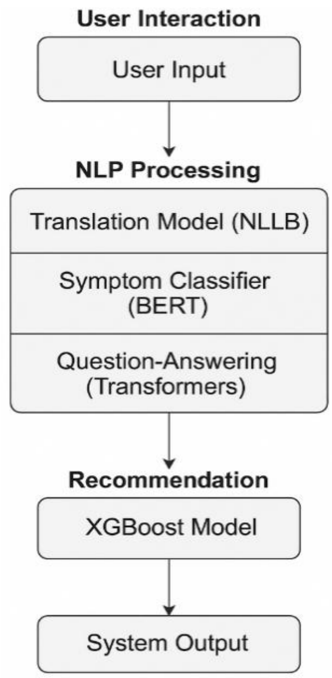

Figure 1. Hybrid VHA architecture

3.2. Dataset description

Two complementary datasets were employed to develop and evaluate the proposed CIMAS HealthMate framework, reflecting the distinct requirements of natural language understanding and structured clinical decision-making. The first dataset comprised a synthetic multilingual symptom corpus containing more than 1,000 annotated symptom–label pairs in English and Shona. The corpus was constructed to support the training of the NLP components, including intent classification and medical entity recognition. Each sample consisted of a free-text patient query annotated with symptom categories, clinical entities, and corresponding intent labels, enabling the model to learn multilingual symptom interpretation and contextual understanding.

The second dataset consisted of an anonymized EHR-style dataset containing approximately 23,000 structured patient records used to train the XGBoost-based triage recommendation model. Each record included clinically relevant attributes such as patient demographics (age and gender), reported symptoms, symptom duration, symptom severity, current medications, pre-existing medical conditions, warning signs (red-flag symptoms), and the corresponding clinician-defined triage recommendation. Based on these structured clinical features, the model was trained to classify patients into one of three triage categories: home_remedy, wait_and_monitor, or see_doctor. The combination of multilingual unstructured text data and structured clinical records enabled the proposed framework to integrate contextual language understanding with interpretable clinical decision-making. This dual-dataset strategy improved the robustness of symptom interpretation while allowing the triage model to learn clinically meaningful decision patterns from structured healthcare data.

3.3. NLP module

The NLP module is responsible for converting free-text patient narratives into structured clinical information that can be processed by the triage recommendation model. To support multilingual interactions, the module integrates machine translation, intent classification, and medical entity recognition within a unified processing pipeline. The workflow begins with language identification, where the system detects whether the user's input is written in English or Shona. When a Shona query is detected, the text is translated into English using the offline (NLLB-200-distilled-600M) model. Offline translation enables the system to operate without internet connectivity while preserving patient privacy and reducing inference latency in bandwidth-constrained environments.

The translated text is subsequently processed using a BERT-based language understanding model to identify the user's clinical intent and extract medically relevant entities. The intent classification component categorizes the patient's request into predefined healthcare intents, such as symptom reporting or health consultation. Simultaneously, the medical entity recognition component identifies key clinical concepts, including reported symptoms, symptom duration, severity, current medications, pre-existing conditions, and other contextual information embedded in the patient's narrative. The extracted entities are normalized and transformed into structured feature vectors through categorical encoding and numerical representation. These feature vectors provide a standardized representation of the patient's clinical profile and serve as the input to the XGBoost-based triage recommendation model. By separating language understanding from clinical decision-making, the proposed architecture improves system interpretability, facilitates modular model updates, and enables independent optimization of each component. The combination of multilingual translation and transformer-based contextual language understanding allows the proposed NLP module to accurately interpret patient narratives across languages while preserving clinically relevant information required for evidence-based triage recommendations. This design enhances the robustness and scalability of the overall framework, making it suitable for multilingual healthcare applications in resource-constrained environments.

3.4. XGBoost-based triage model

The final clinical recommendation is generated using an XGBoost classifier, which was selected because of its high predictive accuracy, computational efficiency, and inherent interpretability when applied to structured clinical data. Compared with conventional machine learning algorithms, XGBoost effectively captures complex nonlinear relationships among clinical variables while incorporating regularization techniques that reduce overfitting and improve model generalization.

The input to the XGBoost model consists of structured feature vectors produced by the NLP module. These features include patient demographics (age and gender), reported symptoms, symptom duration, symptom severity, current medications, pre-existing medical conditions, identified red-flag symptoms, and other clinically relevant attributes extracted from the patient conversation. Prior to model training, categorical variables were encoded into numerical representations, and missing values were handled using the native missing-value processing capability of XGBoost.

The model was formulated as a multiclass classification problem to predict one of three evidence-based triage recommendations: home_remedy, wait_and_monitor, or see_doctor. Rather than providing a definitive medical diagnosis, the classifier estimates the most appropriate level of clinical attention based on the available patient information. This approach supports early healthcare guidance while ensuring that final diagnostic decisions remain under the supervision of qualified healthcare professionals.

During inference, the trained XGBoost model evaluates the structured clinical feature vector and assigns a probability score to each triage category. The recommendation with the highest posterior probability is selected as the system output and subsequently translated into a user-friendly response. In addition, the tree-based architecture enables feature importance analysis, allowing the contribution of individual clinical variables to be examined. This level of transparency enhances the interpretability of the decision-making process and supports greater trust in AI-assisted healthcare applications.

The integration of XGBoost with the transformer-based NLP module combines the strengths of contextual language understanding and explainable machine learning. While the NLP component converts unstructured patient narratives into structured clinical information, the XGBoost classifier performs robust evidence-based triage using clinically meaningful features. This hybrid design provides an effective balance between predictive performance, computational efficiency, and interpretability, making it particularly suitable for multilingual, privacy-sensitive, and resource-constrained healthcare environments.

3.5. Model training and evaluation

The proposed hybrid framework was trained by independently optimizing the NLP components and the XGBoost-based triage model before integrating them into a unified inference pipeline. This modular training strategy allowed each component to be optimized according to its specific learning objective while maintaining flexibility for future model updates. For the NLP module, the multilingual symptom corpus was used to train the intent classification and medical entity recognition models. The translation component employed the pre-trained NLLB-200-distilled-600M model without additional fine-tuning, whereas the BERT-based language understanding model was trained to recognize healthcare intents and extract clinically relevant entities from patient narratives. The extracted information was subsequently transformed into structured feature vectors that served as input to the triage recommendation model. The XGBoost classifier was trained using the anonymized EHR-style dataset containing approximately 23,000 structured patient records. Prior to training, categorical variables were encoded into numerical representations, while missing values were handled using the native missing-value capability of XGBoost. The model was formulated as a three-class classification problem to predict one of the predefined triage recommendations: home_remedy, wait_and_monitor, or see_doctor. To improve model robustness and reduce overfitting, hyperparameters were optimized using cross-validation, and the final model was selected based on its validation performance.

Model performance was evaluated using standard multiclass classification metrics, including accuracy, precision, recall, and F1-score, which collectively assess predictive performance across all triage categories. In addition, translation quality was measured using bidirectional English–Shona translation accuracy, while the NLP module was evaluated based on symptom extraction accuracy. The overall effectiveness of the proposed framework was assessed through end-to-end triage prediction, measuring the system's ability to transform multilingual patient narratives into clinically appropriate recommendations. The modular evaluation strategy enabled the performance of each subsystem (translation, language understanding, and clinical decision-making) to be assessed independently before evaluating the integrated framework. This comprehensive evaluation provides a more reliable assessment of the proposed hybrid architecture and facilitates the identification of potential performance bottlenecks in multilingual AI-assisted healthcare applications.

3.6. Real-time inference workflow

During deployment, CIMAS HealthMate performs symptom triage through a sequential real-time inference pipeline that integrates multilingual NLP with explainable machine learning. Figure 2 illustrates the overall inference workflow, beginning with patient input and ending with an evidence-based triage recommendation. The workflow starts when a patient submits a symptom description through the conversational interface. The system first detects the input language and, when necessary, translates Shona text into English using the offline NLLB-200 translation model. This preprocessing step standardizes multilingual inputs while ensuring reliable operation without internet connectivity.

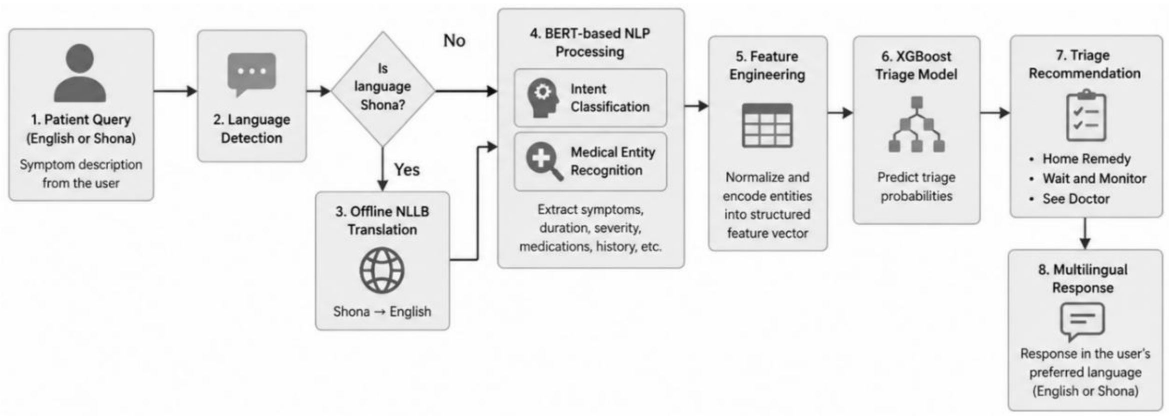

Figure 2. Real-time inference workflow of CIMAS HealthMate

The translated text is then processed by the BERT-based NLP module to identify the user's clinical intent and extract medically relevant entities, including reported symptoms, symptom duration, severity, current medications, pre-existing medical conditions, and other contextual information. These extracted entities are normalized and transformed into a structured feature vector that represents the patient's clinical profile. The structured feature vector is subsequently passed to the trained XGBoost classifier, which evaluates the patient's condition and estimates the probability of each predefined triage category. The class with the highest predicted probability is selected as the final recommendation, corresponding to one of three clinical outcomes: home_remedy, wait_and_monitor, or see_doctor. Rather than providing a medical diagnosis, the system generates evidence-based triage guidance to assist patients in determining the appropriate level of healthcare attention.

Finally, the predicted recommendation is converted into a natural language response and presented to the user in the preferred language. When the original query is submitted in Shona, the recommendation is translated back into Shona using the offline translation module. This bidirectional multilingual workflow ensures linguistic accessibility while preserving the clinical meaning of the recommendation. The modular inference pipeline enables each processing component to operate independently while exchanging standardized structured information. Consequently, the proposed framework achieves low-latency, privacy-preserving, and multilingual symptom triage that is suitable for deployment in low-bandwidth and resource-constrained healthcare environments.

3.7. Implementation environment

The proposed CIMAS HealthMate framework was implemented using Python as the primary development platform, integrating the hugging face transformers, XGBoost, and scikit-learn libraries to support multilingual NLP and explainable machine learning. A Streamlit-based web interface was developed to provide an interactive conversational environment for patient–system communication, enabling real-time symptom assessment and triage recommendations. To ensure efficient deployment and minimize dependence on external services, all trained models were deployed locally. The BERT-based intent classification model, the feature engineering pipeline, and the XGBoost triage classifier were serialized using the joblib framework and reloaded during runtime. Similarly, the NLLB-200-distilled-600M translation model was executed entirely offline, eliminating the need for cloud-based translation services. This deployment strategy improves data privacy, reduces inference latency, and enables reliable operation in environments with limited or intermittent internet connectivity.

The system was designed as a modular processing pipeline in which each component performs an independent task while exchanging standardized structured data with subsequent modules. Such a modular architecture simplifies maintenance, facilitates independent model updates, and allows additional languages or clinical decision models to be incorporated without redesigning the entire framework. Furthermore, the local execution of all processing components supports compliance with privacy-by-design principles by ensuring that sensitive patient information remains within the deployment environment. Overall, the proposed implementation provides a lightweight, scalable, and privacy-preserving platform capable of delivering multilingual AI-assisted symptom triage in real time. These characteristics make CIMAS HealthMate particularly suitable for deployment in low-resource healthcare settings, where limited computational infrastructure, restricted network connectivity, and stringent data privacy requirements present significant operational challenges.

4. RESULTS AND DISCUSSION

This section presents the experimental evaluation and discussion of the proposed CIMAS HealthMate framework. The evaluation aims to assess the effectiveness of the hybrid architecture in supporting multilingual symptom understanding, accurate clinical triage, and real-time healthcare assistance in low-resource environments. Performance is analyzed from multiple perspectives, including user interface usability, translation quality, symptom extraction, triage classification accuracy, computational efficiency, and privacy-preserving deployment.

The proposed framework is evaluated using both qualitative and quantitative analyses. Qualitative evaluation demonstrates the usability of the multilingual conversational interface through representative interaction scenarios in English and Shona, while quantitative evaluation measures the performance of each system component using standard machine learning metrics, including translation accuracy, symptom extraction accuracy, precision, recall, and F1-score. In addition, the real-time inference capability and deployment characteristics of the system are discussed to assess its practical suitability for resource-constrained healthcare settings. The experimental results demonstrate that the integration of transformer-based NLP with an explainable XGBoost classifier provides reliable multilingual symptom interpretation and accurate evidence-based triage recommendations. The following subsections present the system interface, experimental findings, comparative performance analysis, and discussion of the proposed framework's strengths, limitations, and practical implications for AI-assisted healthcare.

4.1. User interface demonstration

The CIMAS HealthMate interface was designed to provide an intuitive and accessible platform for multilingual AI-assisted symptom triage. Emphasis was placed on simplicity, ease of navigation, and patient-centered interaction to ensure that individuals with varying levels of digital literacy can effectively use the system. The interface supports both English and Shona, enabling users to communicate symptoms in their preferred language while maintaining a consistent conversational workflow. Figure 3 presents the authentication and dashboard interface of CIMAS HealthMate. The authentication page provides secure user access through registration and login functions, while the dashboard serves as the central interaction hub, offering access to AI-assisted symptom assessment, conversation history, and system functions. The dashboard also displays multilingual prompts and privacy reminders to encourage user confidence and reinforce responsible handling of sensitive health information.

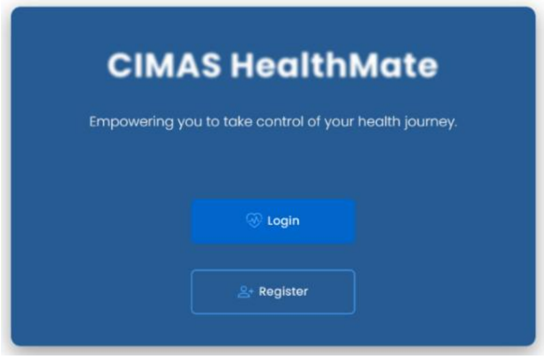

Figure 3. CIMAS HealthMate dashboard with AI diagnoses features

Figure 4 demonstrates representative interactions between users and the VHA in both Shona and English. After receiving a patient's symptom description, the system automatically identifies the input language, performs multilingual NLP, extracts clinically relevant information, and generates an evidence-based triage recommendation. The responses are presented in natural language using culturally appropriate terminology, ensuring that healthcare guidance remains understandable and accessible to users regardless of their language preference.

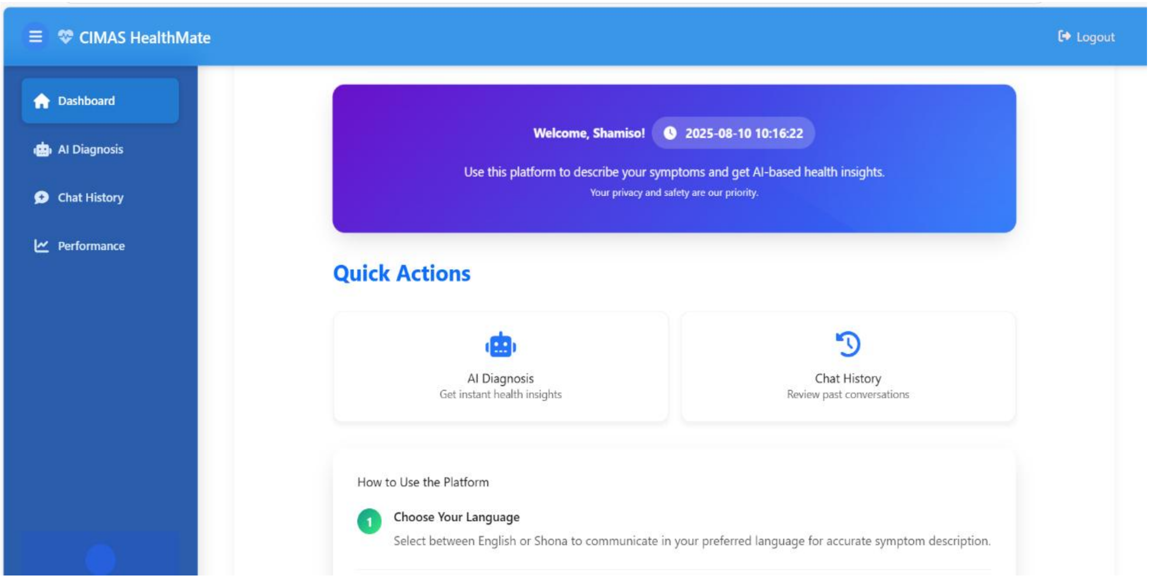

Figure 4. CIMAS HealthMate dashboard

The demonstrated interactions confirm that the proposed framework successfully integrates multilingual communication, contextual language understanding, and explainable clinical decision-making within a unified conversational interface. By providing real-time symptom assessment and user-friendly feedback, the interface enhances accessibility while preserving the transparency and usability required for AI-assisted healthcare applications in low-resource environments.

4.2. Experimental performance evaluation

The performance of the proposed CIMAS HealthMate framework was evaluated by assessing the effectiveness of its individual components and the overall end-to-end triage system. The evaluation focused on multilingual translation, symptom extraction, and clinical triage recommendation. Table 2 summarizes the quantitative performance metrics, while Figure 5 provides a graphical comparison of the translation, symptom extraction, and triage accuracies obtained by the proposed framework.

Table 2. Performance evaluation of the proposed CIMAS HealthMate framework

Task	Performance metric	Result (%)
Shona→English translation	Accuracy	76.5
English→Shona translation	Accuracy	82.2
Symptom extraction	Accuracy	86.6
End-to-end triage prediction	Accuracy	93.3
End-to-end triage prediction	Weighted F1-score	93.3
End-to-end triage prediction	Weighted precision	92.9

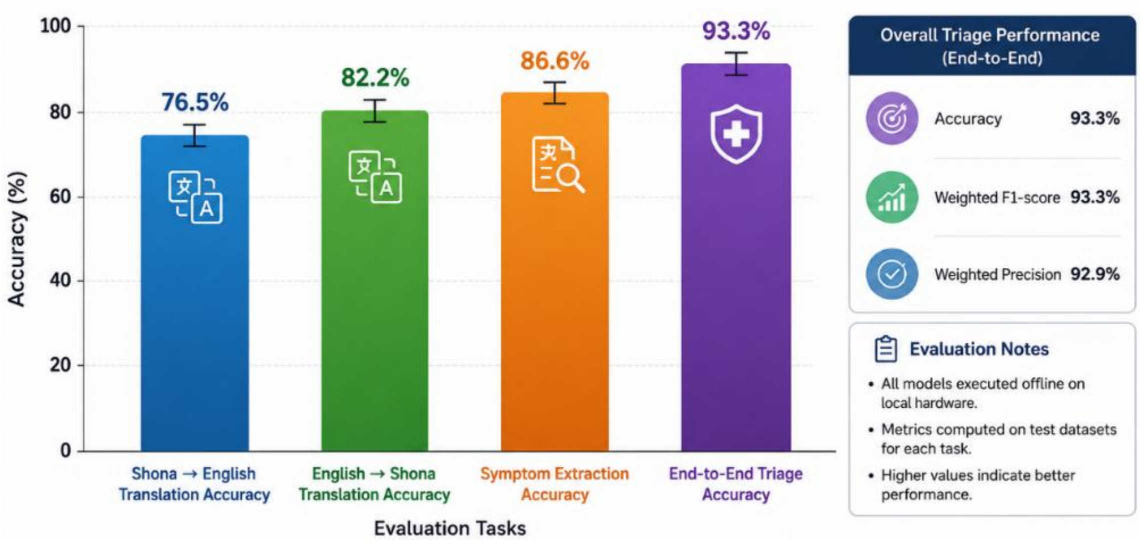

Figure 5. Experimental performance of the proposed CIMAS HealthMate framework, showing translation accuracy, symptom extraction accuracy, and end-to-end triage performance

As illustrated in Figure 5, the hybrid architecture consistently achieved strong performance across all evaluation tasks. The offline NLLB translation model achieved translation accuracies of 76.5% for Shona-to-English and 82.2% for English-to-Shona translation. The BERT-based NLP module attained a symptom extraction accuracy of 86.6%, demonstrating its capability to identify clinically relevant entities from multilingual patient narratives. The highest performance was obtained for end-to-end triage prediction, where the proposed framework achieved 93.3% accuracy with a weighted F1-score of 93.3%, indicating robust and reliable clinical decision support.

The experimental results demonstrate that the proposed hybrid architecture achieves strong overall performance across multilingual healthcare tasks. The offline NLLB translation model successfully supports bidirectional communication between English and Shona, although translation performance is higher for English-to-Shona conversion than for Shona-to-English translation. This performance difference is expected because Shona remains a relatively low-resource language with fewer large-scale parallel corpora available for pre-training and evaluation.

The BERT-based NLP module achieved a symptom extraction accuracy of 86.6%, indicating its capability to accurately identify clinically relevant entities from free-text patient narratives. Effective extraction of symptoms, medications, duration, severity, and contextual clinical information provides reliable structured inputs for the downstream decision-making process. These findings demonstrate the effectiveness of transformer-based contextual representations in interpreting multilingual healthcare conversations.

The highest performance was obtained in the end-to-end triage task, where the proposed framework achieved 93.3% classification accuracy with a weighted F1-score of 93.3%. This result indicates that the integration of multilingual language understanding with an explainable XGBoost classifier enables robust and consistent clinical recommendation across the three predefined triage categories (home_remedy, wait_and_monitor, and see_doctor). The high weighted precision further suggests that the classifier maintains a balanced prediction capability while minimizing incorrect triage recommendations. Overall, the experimental findings confirm that combining transformer-based NLP with XGBoost-based clinical decision-making effectively bridges the gap between unstructured multilingual patient narratives and structured evidence-based triage. The modular design allows each subsystem to optimize its respective task while collectively delivering accurate, interpretable, and privacy-preserving healthcare assistance suitable for deployment in multilingual and resource-constrained environments.

4.3. Comparative analysis

To further evaluate the reliability of the proposed triage model, the classification performance was analyzed using a confusion matrix. Figure 6 illustrates the distribution of correctly and incorrectly classified instances across the three triage categories (home_remedy, wait_and_monitor, and see_doctor). The confusion matrix provides a more detailed assessment of model behavior than overall accuracy by identifying the specific classes in which prediction errors occur.

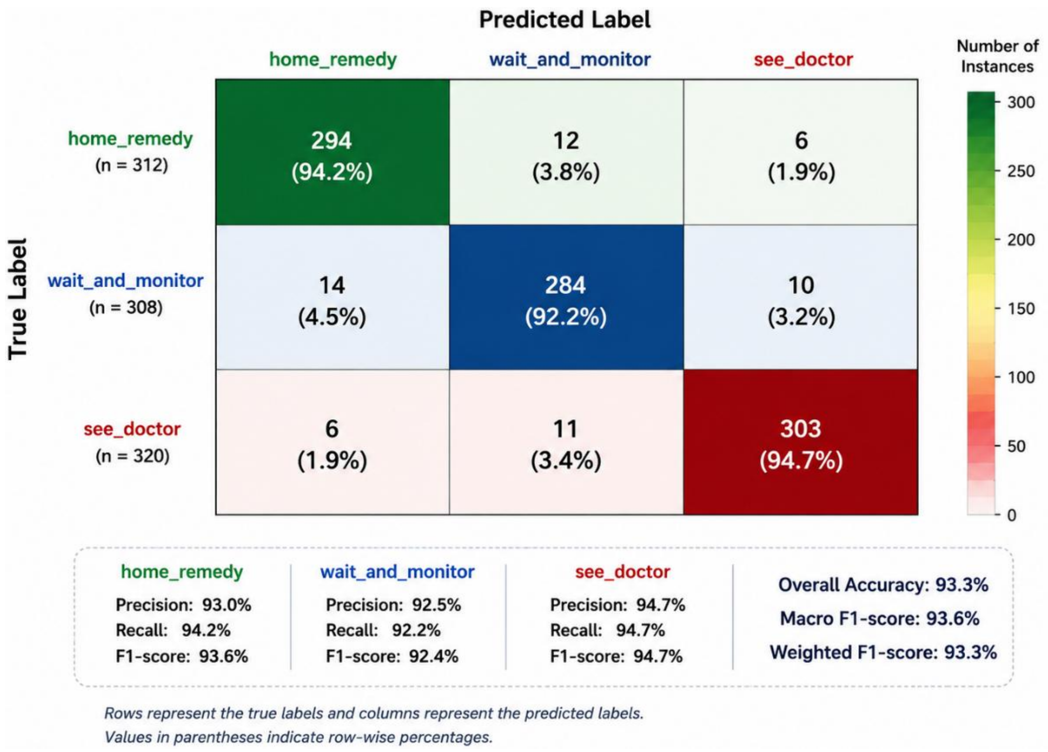

Figure 6. Confusion matrix of the proposed XGBoost-based triage classifier

As shown in Figure 6, most test samples were correctly assigned to their respective triage categories, with only a small number of misclassifications occurring between neighboring clinical recommendations. The relatively small number of off-diagonal elements indicates that the proposed XGBoost classifier effectively distinguishes between different levels of clinical urgency while maintaining balanced prediction performance across all classes. These findings are consistent with the overall classification accuracy of 93.3% and the weighted F1-score of 93.3%, confirming the robustness of the proposed hybrid framework for multilingual symptom triage.

To assess the effectiveness of the proposed hybrid framework, CIMAS HealthMate was compared with representative machine learning and deep learning approaches commonly used for AI-assisted healthcare applications. The comparison focused on multilingual capability, model interpretability, deployment suitability, and clinical triage performance. Table 3 summarizes the qualitative comparison between the proposed framework and representative baseline methods.

Unlike conventional machine learning methods, which rely primarily on structured numerical features, the proposed framework combines multilingual transformer-based language understanding with an explainable gradient-boosting classifier. The BERT-based NLP module enables accurate interpretation of unstructured patient narratives, while the XGBoost classifier performs transparent and robust clinical triage using structured clinical features extracted from the conversation. This complementary design exploits the strengths of both deep learning and traditional machine learning while mitigating their individual limitations.

Compared with end-to-end neural conversational systems, the proposed architecture offers greater interpretability because the clinical recommendation is generated by a tree-based decision model whose feature contributions can be analyzed using feature importance techniques. Furthermore, executing all components locally eliminates dependence on cloud infrastructure, thereby improving privacy protection and ensuring reliable operation in environments with limited internet connectivity.

The experimental evaluation demonstrated that the proposed framework achieved an overall triage accuracy of 93.3% while maintaining multilingual communication through offline English–Shona translation. Although translation accuracy remains lower than triage accuracy due to the linguistic complexity and limited training resources available for Shona, the language understanding module successfully preserved clinically relevant information, enabling the XGBoost classifier to generate accurate evidence-based triage recommendations. These findings indicate that integrating multilingual NLP with explainable machine learning provides an effective balance between predictive performance, interpretability, and deployment practicality for healthcare applications in resource-constrained environments.

Table 3. Comparison of the proposed CIMAS HealthMate framework with representative AI-based healthcare approaches

Method	Multilingual support	Explainability	Offline deployment	Structured clinical reasoning	Suitable for low-resource settings
Logistic regression	Limited	High	Yes	Moderate	Yes
Random forest	Limited	Moderate	Yes	High	Yes
BERT-based NLP	High	Low	Limited	Low	Moderate
XGBoost	No	High	Yes	High	Yes
Proposed CIMAS HealthMate	High (English–Shona)	High	Yes	High	High

4.4. Real-time performance

The real-time performance of CIMAS HealthMate was evaluated to assess its suitability for deployment in low-resource healthcare environments where timely clinical guidance is essential. The evaluation focused on the responsiveness of the complete inference pipeline, including multilingual translation, natural language understanding, structured feature generation, and XGBoost-based triage prediction. Experimental observations showed that the computational inference performed by the AI models required only a few milliseconds for each processing stage after model initialization. The overall user interaction, including symptom collection, contextual follow-up questions, and generation of the final triage recommendation, was typically completed within 45–60 seconds. This interaction time primarily depended on the user’s response speed rather than the computational complexity of the underlying AI models.

The modular pipeline contributed significantly to the system’s computational efficiency. The offline NLLB translation model processed multilingual inputs without requiring communication with external servers, while the BERT-based NLP module efficiently extracted clinically relevant entities from patient narratives. Subsequently, the XGBoost classifier generated triage recommendations with minimal computational overhead due to its lightweight tree-based inference mechanism. Local execution of all processing components further reduced network latency and eliminated dependence on cloud-based services.

The observed response time is considered appropriate for real-time symptom triage applications, where rapid patient guidance is more critical than instantaneous conversational responses. Moreover, the offline deployment strategy improves system reliability in environments with unstable internet connectivity while ensuring continuous service availability and enhanced protection of sensitive patient information. Overall, the experimental results demonstrate that the proposed framework provides efficient real-time multilingual healthcare assistance by combining low-latency inference, privacy-preserving local execution, and explainable clinical decision-making. These characteristics make CIMAS HealthMate well suited for deployment in rural clinics, community health centers, and other resource-constrained healthcare settings where computational resources and network infrastructure may be limited.

4.5. Discussion

The experimental results demonstrate that the proposed hybrid architecture effectively combines the complementary strengths of transformer-based NLP and XGBoost-based clinical decision-making.

The multilingual NLP module successfully transformed unstructured patient narratives into structured clinical information, while the XGBoost classifier produced accurate and interpretable triage recommendations. This separation of language understanding and clinical reasoning contributed to the overall robustness of the framework and enabled each component to be independently optimized for its respective task [2], [3].

To further improve model transparency, shapley additive explanations (SHAP) was employed to interpret the contribution of each clinical feature to the triage decision. Figure 7 presents the SHAP summary plot, which ranks features according to their overall influence on the XGBoost classifier. Symptom severity, warning signs, symptom duration, and patient age were identified as the most influential predictors, indicating that the model primarily relies on clinically relevant information rather than demographic variables alone. The distribution of SHAP values further illustrates how high and low feature values affect the predicted triage category, thereby improving the interpretability and trustworthiness of the proposed clinical decision-support system [23].

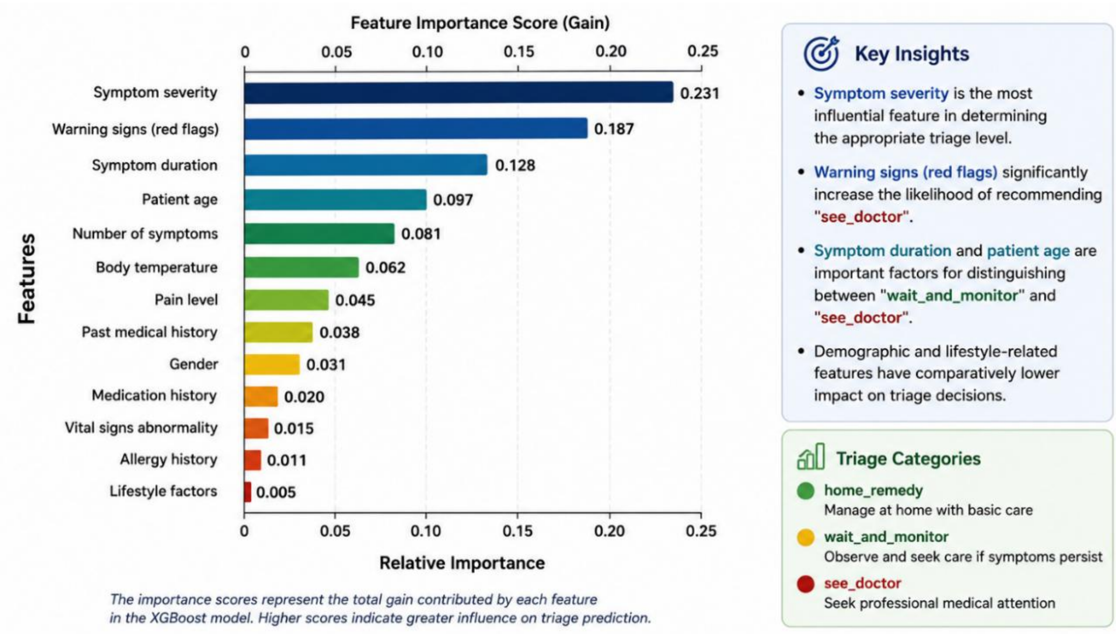

Figure 7. SHAP summary plot (preferred) or XGBoost feature importance

The end-to-end triage accuracy of 93.3% indicates that integrating contextual language understanding with structured machine learning provides reliable evidence-based healthcare guidance. Compared with end-to-end neural conversational models, the proposed architecture offers greater interpretability because the final recommendation is generated using a tree-based classifier whose feature contributions can be examined through feature importance analysis. Such transparency is particularly important in healthcare applications, where clinicians and patients require understandable AI-assisted recommendations rather than opaque predictions [19].

Although the translation module achieved satisfactory multilingual performance, the accuracy for Shona-to-English translation (76.5%) remained lower than that for English-to-Shona translation (82.2%). This performance difference can largely be attributed to the limited availability of high-quality parallel corpora and linguistic resources for Shona compared with English. Nevertheless, the symptom extraction module maintained an accuracy of 86.6%, suggesting that the translation process preserved sufficient clinical information for reliable downstream triage. These findings indicate that robust symptom interpretation can still be achieved despite moderate translation accuracy, demonstrating the resilience of the proposed hybrid pipeline [18].

From a deployment perspective, the offline implementation represents a significant advantage for healthcare delivery in low-resource settings. By executing translation, language understanding, and clinical decision-making locally, the system minimizes dependence on cloud infrastructure, reduces communication latency, and strengthens patient data privacy. These characteristics make the framework particularly suitable for rural healthcare facilities and community clinics where internet connectivity may be unreliable or unavailable [15], [25].

Despite these encouraging results, several limitations should be acknowledged. The multilingual NLP module currently supports only English and Shona, and the evaluation was conducted using a combination of synthetic multilingual symptom data and anonymized structured clinical records. Consequently, the reported performance may not fully represent the linguistic diversity and variability encountered in routine clinical practice. Furthermore, the proposed system is designed exclusively for symptom triage rather than disease diagnosis and should therefore be considered a decision-support tool intended to complement, rather than replace, professional clinical judgment.

Overall, the experimental findings demonstrate that integrating multilingual transformer models with an explainable XGBoost classifier provides an effective balance between predictive accuracy, interpretability, computational efficiency, and deployment practicality. The proposed framework offers a promising approach for expanding equitable access to AI-assisted healthcare while maintaining transparency, privacy, and operational reliability in multilingual and resource-constrained healthcare environments [2], [11].

4.6. Privacy, ethics, and clinical implications

The deployment of AI-assisted healthcare systems requires careful consideration of patient privacy, ethical responsibility, and clinical safety. Accordingly, CIMAS HealthMate was designed following privacy-by-design and human-centered AI principles to ensure that the system provides safe, transparent, and trustworthy symptom triage while protecting sensitive patient information. From a privacy perspective, all data processing is performed locally within the deployment environment. The offline execution of the NLLB translation model, BERT-based NLP module, and XGBoost classifier eliminates the need to transmit patient information to external cloud services, thereby reducing the risk of unauthorized data disclosure. Furthermore, the system does not permanently store personally identifiable health information after each consultation session. Only anonymized operational metadata, such as processing time and model performance statistics, may be retained for system monitoring and quality improvement. This architecture supports compliance with internationally recognized data protection frameworks, including the general data protection regulation (GDPR), the health insurance portability and accountability act (HIPAA), and Zimbabwe's cyber and data protection act (CDPA).

Ethically, the proposed framework is designed exclusively as a clinical decision support system for symptom triage rather than as an autonomous diagnostic tool. The system provides one of three evidence-based recommendations: home_remedy, wait_and_monitor, or see_doctor, without attempting to diagnose diseases or prescribe treatments. Consequently, clinical responsibility remains with qualified healthcare professionals, while the AI system serves to assist patients in determining the appropriate level of medical attention. This human-in-the-loop approach reduces the risk of overreliance on automated recommendations and promotes responsible integration of AI into healthcare practice.

From a clinical perspective, the multilingual and offline capabilities of CIMAS HealthMate have important implications for healthcare delivery in resource-constrained settings. By supporting both English and Shona, the system improves communication with patients who may have limited English proficiency, thereby enhanced accessibility and reducing language-related barriers to healthcare. In addition, offline deployment enables reliable operation in rural and underserved communities where internet connectivity is limited or unavailable. These characteristics improve the practicality of AI-assisted triage while strengthening patient trust through transparent, explainable, and privacy-preserving decision support.

Overall, the proposed framework demonstrates that responsible AI design extends beyond predictive performance to include ethical deployment, regulatory compliance, clinical transparency, and equitable access to healthcare. Integrating these considerations into the system architecture enhances the feasibility of deploying multilingual VHAs in real-world healthcare environments while maintaining patient safety, data confidentiality, and professional clinical oversight.

4.7. Limitations and future work

Although the proposed CIMAS HealthMate framework demonstrates promising performance for multilingual AI-assisted symptom triage, several limitations should be acknowledged. First, the NLP module currently supports only English and Shona, limiting its applicability to other indigenous African languages. Expanding multilingual coverage would improve accessibility across diverse linguistic communities. Second, the symptom classification model was trained using a synthetic multilingual dataset, while the triage recommendation model relied on anonymized structured EHR-style data. Although these datasets facilitated model development and evaluation, additional validation using large-scale real-world clinical conversations is necessary to further assess the system's robustness and generalizability.

Third, the current evaluation was conducted in a controlled experimental environment and did not include prospective clinical trials or usability studies involving healthcare professionals and patients. Consequently, the reported performance primarily reflects technical effectiveness rather than real-world clinical adoption. Furthermore, the proposed framework is designed exclusively for symptom triage and preliminary healthcare guidance. It does not provide disease diagnosis, treatment recommendations, or emergency medical decision-making, and therefore should be considered a clinical decision support tool that complements rather than replaces professional medical judgment.

Future research will focus on improving both the technical capabilities and clinical applicability of the framework. Planned enhancements include fine-tuning multilingual transformer models using region-specific healthcare conversations to improve symptom understanding and translation quality for African languages. Privacy-preserving federated learning will also be investigated to enable continuous model improvement without centralized collection of sensitive patient data. To increase accessibility, future versions will support deployment through low-bandwidth communication platforms such as WhatsApp and short message service (SMS), allowing wider adoption in remote and underserved communities. In addition, integration with hospital information systems through health level seven international fast healthcare interoperability resources (HL7 FHIR)-compliant EHR interfaces will facilitate interoperability with existing clinical workflows. Finally, large-scale clinical validation involving healthcare practitioners, multilingual patient populations, and multiple healthcare institutions will be conducted to evaluate usability, clinical effectiveness, and long-term deployment in routine healthcare settings. Overall, these future developments are expected to enhance the scalability, multilingual capability, clinical reliability, and practical deployment of CIMAS HealthMate, thereby strengthening its potential as an explainable and privacy-preserving AI-assisted healthcare solution for resource-constrained environments.

5. CONCLUSION

This study presented CIMAS HealthMate, a multilingual VHA designed to support AI-assisted symptom triage in resource-constrained healthcare environments. The proposed framework integrates offline NLLB-based multilingual translation, BERT-based NLP, and an explainable XGBoost classifier to provide evidence-based triage recommendations while preserving user privacy and enabling deployment without continuous internet connectivity. By combining contextual language understanding with structured machine learning, the system effectively supports English–Shona communication and transparent clinical decision support. Experimental evaluation demonstrated the effectiveness of the proposed hybrid architecture, achieving translation accuracies of 76.5% (Shona-to-English) and 82.2% (English-to-Shona), a symptom extraction accuracy of 86.6%, and an end-to-end triage accuracy and weighted F1-score of 93.3%. These results indicate that the framework provides reliable multilingual symptom assessment while maintaining computational efficiency, explainability, and suitability for low-bandwidth healthcare settings. Beyond its technical performance, CIMAS HealthMate demonstrates the feasibility of integrating multilingual AI with interpretable machine learning to improve access to preliminary healthcare guidance in underserved communities. The privacy-preserving, offline architecture further supports deployment in environments where network infrastructure and clinical resources are limited. Future work will focus on expanding support for additional African languages, validating the system using real-world clinical datasets and prospective user studies, integrating HL7 FHIR-compliant EHR systems, and investigating federated learning to enable continuous model improvement while preserving patient privacy.

ACKNOWLEDGEMENTS

The authors would like to express their sincere appreciation to all individuals and institutions whose support, encouragement, and technical assistance contributed to the successful completion of this research. The authors also acknowledge the valuable feedback and constructive comments provided during the development of this work, which helped improve the quality of the study.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOUR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Shamiso Simango	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Mainford Mutandavari	✓	✓		✓		✓		✓		✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data supporting the findings of this study can be obtained by contacting the corresponding author, [SS]. This data includes information which may compromise privacy of the research participants and therefore not publicly accessible because of specific restrictions.

REFERENCES

[1] A. Vaswani *et al.*, "Attention is all you need," *arXiv*:1706.03762, 2017.
[2] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," *arXiv*:1810.04805, 2018.
[4] P. N. Gosavi, "Comparative study of transformer models for text classification in healthcare," M.S. thesis, School of Computing, National College of Ireland, Dublin, Ireland, 2024.
[5] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep learning-based text classification: a comprehensive review," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–40, Apr. 2022, doi: 10.1145/3439726.
[6] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–35, Sep. 2023, doi: 10.1145/3560815.
[7] A. B. Abacha, Y. Mrabet, M. Sharp, T. R. Goodwin, S. E. Shooshan, and D. Demner-Fushman, "Bridging the gap between consumers' medication questions and trusted answers," in *MEDINFO 2019: Health and Wellbeing e-Networks for All*, 2019, pp. 25–29, doi: 10.3233/SHTI190176.
[8] L. Abou *et al.*, "Gait and balance assessments using smartphone applications in Parkinson's disease: a systematic review," *Journal of Medical Systems*, vol. 45, no. 9, p. 87, Sep. 2021, doi: 10.1007/s10916-021-01760-5.
[9] M. Rashed, M. I. Hossain, A. Mahdi, and G. Mustofa, "Hybrid machine learning models for accurate type 2 diabetes mellitus prediction using a stacking classifier and a meta-model approach," *Cureus Journal of Computer Science*, vol. 9, pp. 1–22, Mar. 2025, doi: 10.7759/s44389-025-03135-0.
[10] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, pp. 76516–76531, 2020, doi: 10.1109/ACCESS.2020.2989857.
[11] G. M. S. Himel, M. S. Hasan, U. S. Salsabil, and M. M. Islam, "MedLingua: a conceptual framework for a multilingual medical conversational agent," *MethodsX*, vol. 12, p. 102614, Jun. 2024, doi: 10.1016/j.mex.2024.102614.
[12] H. Wang, S. Ma, L. Dong, S. Huang, D. Zhang, and F. Wei, "DeepNet: scaling transformers to 1,000 layers," *arXiv*:2203.00555, 2022.
[13] S. Ruder, "Neural transfer learning for natural language processing," Ph.D. dissertation, National University of Ireland, Galway, Ireland, 2019.
[14] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 328–339, doi: 10.18653/v1/P18-1031.
[15] R. Stenhouse *et al.*, "Developing a COVID-19-focused mHealth system in a low-resource setting during the COVID-19 pandemic: challenges and opportunities," *Frontiers in Digital Health*, vol. 7, p. 1543828, Jun. 2025, doi: 10.3389/fdgth.2025.1543828.
[16] K. S. Kalyan, A. Rajasekharan, and S. Sangeetha, "AMMU: a survey of transformer-based biomedical pretrained language models," *Journal of Biomedical Informatics*, vol. 126, p. 103982, Feb. 2022, doi: 10.1016/j.jbi.2021.103982.
[17] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: a lite BERT for self-supervised learning of language representations," *arXiv:*1909.11942, 2019.
[18] A. Conneau *et al.*, "Unsupervised cross-lingual representation learning at scale," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 8440–8451, doi: 10.18653/v1/2020.acl-main.747.
[19] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: can language models be too big?" in *FAccT 2021-Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Mar. 2021, pp. 610–623, doi: 10.1145/3442188.3445922.
[20] K. P. Murphy, *Probabilistic machine learning: an introduction*. Cambridge, MA: MIT Press, 2022.
[21] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv:*2010.16061, 2020.
[22] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," *arXiv:*1811.12808, 2018.

[23] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[24] R. A. Calvo and S. D'Mello, "Affect detection: an interdisciplinary review of models, methods, and their applications," *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, Jan. 2010, doi: 10.1109/T-AFFC.2010.1.

[25] World Health Organization, "Digital health for the end TB strategy: an agenda for action," *WHO*. Accessed Feb. 18, 2025. [Online]. Available: https://www.who.int/publications/i/item/WHO-HTM-TB-2015.21

BIOGRAPHIES OF AUTHORS

Shamiso Simango is an M.Tech. student in Information Technology at Harare Institute of Technology (HIT), Zimbabwe. She holds a Bachelor of Science degree in Applied Information Technology, specializing in Network Engineering from the University of Zimbabwe, and a Diploma in Applied Information Technology specializing in Network Engineering from the University of Zimbabwe. She has obtained a certificate in digital transformation through ICT from the Japan International Cooperation Agency and has also obtained a certificate in cybersecurity from Huawei Technologies (China). She has 8 years of experience in network engineering, telecommunications, and ICT security audits. Her research interests cover topics that relate to network engineering, cybersecurity, internet of things, artificial intelligence, and data analytics. She can be contacted at email: shamisosimaz@gmail.com.

Mainford Mutandavari is a Ph.D. scholar at SRMIST University, India, and a lecturer and Postgraduate Studies Coordinator at the Harare Institute of Technology (HIT), Zimbabwe. With advanced degrees in Computer Science and Strategy and Innovation, his research spans data analytics, cybersecurity, IoT, AI, and cloud computing. He is a member of HIT's Cybersecurity and AI research groups and actively contributes to national ICT standards through the Standards Association of Zimbabwe. Mainford has published widely on topics such as data loss prevention systems, digital learning infrastructure, and e-health security. His work bridges academic research with industry applications, focusing on practical digital solutions for education, telecommunications, and healthcare in Zimbabwe. He is also involved in curriculum development, postgraduate supervision, and building academic-industry partnerships**.** He can be contacted at email: mmutanda.vari@gmail.com.