Received: 06 SEPT 2024 | Accepted: 31 MAR 2025 | Published Online: 05 APR 2025
Recoletos Multidisciplinary Research Journal Vol. 1 Special Issue (2025)
DOI: https://doi.org/10.32871/rmrj25si.t2407

recoletos MULTIDISCIPLINARY RESEARCH JOURNAL RMRJ

Original Article

Pitik: A Cebuano-Binisaya Intent-Based Chatbot for Cardiovascular Disease Patient Profiling and Risk Factor Recommendations

Joseph G. Cedeño[1], Andrew E. Manteza[1], Nicole C. Nacar[1], Merhamdin P. Umbukan[1], Cherrie G. Muaña[2], Ma. Juliet Vasay-Cruz[3], Ceasar Ian P. Benablo[1], Kristine Mae M. Adlaon[4*]

[1] College of Computer Studies, University of the Immaculate Conception, Davao City, Philippines

[2] College of Pharmacy and Chemistry, University of the Immaculate Conception, Davao City, Philippines

[3] College of Arts and Humanities, University of the Immaculate Conception, Davao City, Philippines

[4] Mindanao Natural Language Processing R & D Laboratory, College of Computer Studies, University of the Immaculate Conception, Davao City, Philippines

*Correspondence: kadlaon@uic.edu.ph

Abstract

Background: *Cardiovascular diseases (CVDs) remain the leading cause of death in the Philippines, affecting one in six Filipinos and accounting for 20% of all deaths. Despite the existence of community-based healthcare programs, patient profiling continues to be done manually, resulting in inefficiencies in cardiovascular risk assessment. To address this, Pitik, a Cebuano-Binisaya intent-based chatbot, was developed to streamline cardiovascular risk profiling and data collection, particularly in underserved areas.*
Methods: *This study collaboratively employed Action Research to refine Pitik through three software development iterations. The chatbot integrated the Diag-Ex framework alongside Pre-Intent and Post-Intent Matching algorithms. Gricean Maxims guided its conversational design to enhance communication accuracy and user interaction quality.*
Results: *The iterative development process significantly improved Pitik's accuracy, reduced communication errors, and increased user engagement. Evaluations demonstrated the chatbot's effectiveness in processing user inputs and providing structured cardiovascular risk assessments. These improvements highlight Pitik's growing capability in delivering accessible and reliable health information.*
Conclusion: *Pitik presents a scalable and linguistically inclusive AI solution for cardiovascular risk assessment within Cebuano-Binisaya-speaking communities. The study underscores the potential of AI-driven chatbots to enhance community-based patient profiling, reduce manual workloads, and improve healthcare access in rural areas. Future work will involve expanding Pitik's features and evaluating its real-world impact in broader healthcare contexts.*

Keywords

patient profiling, Cardiovascular Disease Risk Factor Assessment, community-based healthcare, Gricean Maxims, medical chatbot

INTRODUCTION

Cardiovascular diseases (CVDs) remain a significant public health challenge in the Philippines, consistently ranking as the leading cause of mortality (Cacciata et al., 2021). These diseases account for approximately 20%

of all deaths and 35% of premature deaths in the country, affecting one in every six Filipinos. High blood pressure, high cholesterol, and smoking are key risk factors, along with other conditions and lifestyle choices such as diabetes, obesity, and overweight. To address these issues outside Metro Manila, one of the initiatives of healthcare professionals is conducting community outreach programs in rural and underserved areas. These initiatives involve profiling residents and providing essential healthcare advice, aiming to mitigate the impact of cardiovascular-related diseases in these communities (Reyes et al., 2023). Depending on the institution, community-based profiling and diagnosing patients involves localized digital systems and sometimes physical records. Health records between institutions are encoded physically through paper and are not standardized between institutions (Evans, 2016; Menachemi, 2011). In many cases, health records are maintained using localized digital systems, often with heavy reliance on physical records, especially in less developed areas. The inconsistency in how patient information is recorded and stored across different institutions leads to significant discrepancies, making it difficult to transfer patient data effectively. This data fragmentation hinders compiling a comprehensive medical history, which is crucial for accurate diagnosis and treatment. While some healthcare providers may have advanced electronic health records (EHR) systems, others may still depend on outdated or manual methods. This disparity complicates data sharing and limits the overall efficiency and accuracy of patient profiling (Quinn et al., 2019; Casey et al., 2016; Iyanna et al, 2022).

The Philippines offers various healthcare applications like KonsultaMD, which provides 24/7 access to licensed doctors via telehealth (Noceda et al., 2023). While these technologies expand healthcare access, they lack crucial elements. None are exclusively available in Cebuano-Binisaya, a language spoken by millions, particularly in rural areas where access to healthcare is already limited. This gap is significant, as language barriers can hinder effective communication and understanding, especially in areas with high rates of cardiovascular diseases. Moreover, no existing Cebuano-Binisaya healthcare datasets, make creating targeted digital tools for these communities challenging. Developing a Cebuano-Binisaya chatbot is essential to bridge these gaps, providing culturally and linguistically appropriate healthcare support, and ensuring that vital health information and services are accessible to those most in need.

To bridge these gaps and bring critical healthcare support to those who need it most, we developed Pitik—a web application designed to speak directly to the heart of the Cebuano-Binisaya community. Pitik is more than just a chatbot; it is a lifeline that helps users navigate their cardiovascular health concerns in their native language, ensuring that no one is left behind. By focusing on underserved populations in remote areas, Pitik aims to bring the benefits of modern healthcare technology to those who have been overlooked for too long. Unlike existing telemedicine platforms, Pitik bridges gaps in healthcare accessibility by digitizing data collection and enabling culturally and linguistically appropriate interactions. Many healthcare professionals such as pharmacists, particularly those more fluent in English and Filipino, face challenges in effectively communicating with patients, as the assessment tools they use are primarily in English. This language barrier has led to misunderstandings, incomplete data collection, and reduced patient adherence to medical advice. By engaging with patients in their native language, pharmacists can ensure clearer communication (Tan et al., 2024), more accurate health data collection, and stronger patient trust, ultimately improving cardiovascular health outcomes in underserved areas.

There are several applications with functionalities like Pitik, each offering unique approaches. For instance, Diabot is a predictive medical chatbot that provides users with general disease and diabetes predictions, encouraging proactive measures such as lifestyle changes and medication adjustments (Sarma et al., 2019). Diabot employs ensemble learning to deliver accurate diagnoses, utilizing datasets such as a general health dataset for disease prediction and the Pima Indian diabetes dataset for diabetes prediction. Other chatbots collect personal and medical details, which are then analyzed using algorithms like Support Vector Machine to identify specific illnesses and suggest treatment options. Additionally, some chatbots follow a linear design, progressing from symptom extraction to mapping and diagnosis, with severe conditions triggering a referral to a doctor, who is then provided with relevant patient details from the database.

Data availability is crucial for developing technology-based solutions, particularly those that heavily rely on data for their functionality. Moreover, the Cebuano-Binisaya language lacks digitized datasets related to heart disease diagnosis or patient profiling. This absence may be attributed to the challenges in translating medical terminology into Cebuano-Binisaya and the potential under-exploration or unavailability of the language's implementation in the medical field, particularly in digital formats accessible to the public. Also, the existing data in this domain are predominantly standardized forms, necessitating the construction of a new dataset from the ground up. This process introduces various challenges, including sourcing the data, determining its scope, and ensuring thorough data cleaning and refinement. Equally important in developing a chatbot is its internal structure. How responses are formulated can significantly influence users' perceptions, interpretation of the information, and subsequent actions. Research indicates that effective advice depends not only on the content but also on the presentation of that content. Like how ChatGPT translations require human discernment for cultural and linguistic accuracy (Liwanag et al., 2024), Pitik also necessitates ongoing refinements and validation from healthcare professionals to ensure that patient profiling and cardiovascular risk assessments are effectively communicated in Cebuano-Binisaya. This process highlights the importance of evaluating chatbot responses in terms of both content and structure to ensure successful and meaningful communication.

Considering these challenges, the primary objective of this study is to develop Pitik, a chatbot designed for heart disease-related inquiries in the Cebuano-Binisaya language. The project aims to achieve this by building a schema-guided dialogue dataset specifically for heart disease-related illnesses in Cebuano-Binisaya, addressing issues related to the internal structure of the chatbot, and discovering a conversational scheme that aligns with Cebuano-Binisaya's conversational flow. Furthermore, the study seeks to employ intrinsic and extrinsic evaluation methods to assess Pitik's performance based on the newly developed conversational scheme, ensuring that the chatbot effectively communicates and meets user expectations.

METHODS

The development of Pitik adhered to an Iterative Software Development process, allowing for continuous enhancements and refinements. As illustrated in Figure 1, this process facilitated systematic progress through multiple stages. In the first iteration, the team collaborated with healthcare professionals and reviewed existing literature to gather comprehensive requirements. This informed the design of a user-friendly interaction medium in Cebuano, ensuring cultural and linguistic relevance. The initial implementation of the chatbot was achieved using Google's DialogFlow, based on a risk assessment form provided by healthcare experts. Early testing underscored the need for additional data and further expert input, highlighting areas for significant improvement. These insights drove subsequent iterations, allowing Pitik to evolve into a more robust and effective tool for cardiovascular risk assessment and patient profiling.

In the second iteration, the design was meticulously refined to align more closely with Gricean Maxims, aiming to minimize errors and reduce user confusion. To further enhance the conversational structure, suggestion chips were introduced, providing users with clear response options and streamlining interactions. The testing phase incorporated valuable feedback from both users and linguistic experts, allowing for a comprehensive review of the chatbot's conversation flow. This rigorous evaluation led to targeted improvements, ensuring a more intuitive and effective user experience.

The third iteration concentrated on optimizing the conversational flow to accommodate the nuances of the Cebuano-Binisaya language better, addressing challenges identified in earlier stages. This phase involved refining the chatbot's ability to handle a broader spectrum of user inputs, enhancing its flexibility and responsiveness. The evaluation process was rigorous, employing the Analytic Hierarchy Process (AHP) method to assess overall quality. Pitik underwent both intrinsic and extrinsic evaluations: intrinsically, its thoroughness in assessing cardiovascular health was scrutinized, while extrinsically, it was evaluated based on user feedback from prior iterations. These assessments ensured that Pitik met the technical requirements and resonated with its target users, enhancing its effectiveness as a healthcare tool.

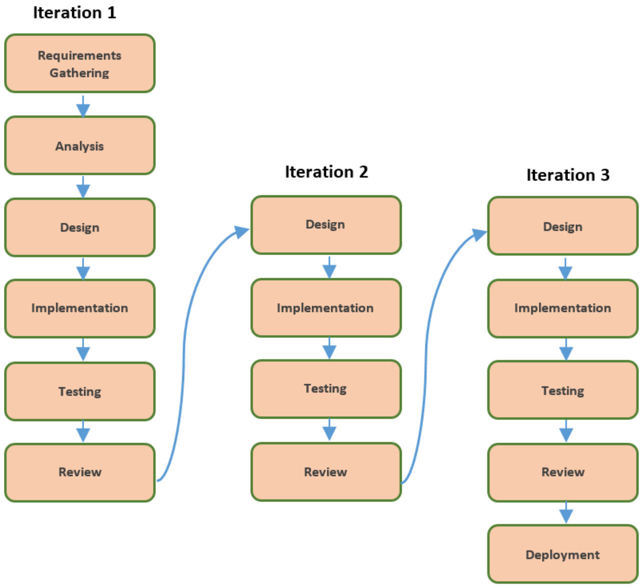

Figure 1. *Iterative Process Model*

The researchers used of Grice's Maxims of Conversation to evaluate the acceptability of the user input which plays a vital role in Pitik's success or failure. Grice's maxims of conversation were chosen as a framework since they serve as a guide in constructing and evaluating the design of our conversational flow. The philosopher Paul Grice (1975) proposed four conversational maxims that explain the link between utterances and what is understood from them. It is based on his cooperative principle which is pragmatic in its approach and is so-called because listeners and speakers must speak cooperatively and mutually accept one another to be understood in a particular way. Linguist Paul Grice proposes four main maxims to describe principles that people intuitively follow to guide their conversations. This work is guided by this principle to have effective communication with the users as shown in Figure 2. After the testing phase, the user's answers were evaluated using this study to see if the chatbot and user had established successful communication. Violations of the Gricean Maxim involve not following the designated rules below; violations may overlap where a sentence or phrase can violate multiple maxims.

Gricean Maxims:

1. The maxim of quantity - where one tries to be as informative as one possibly can, and gives as much information as is needed, and no more.
2. The maxim of quality - where one tries to be truthful and does not give information that is false or that is not supported by evidence
3. The maxim of relation - where one tries to be relevant and says things that are pertinent to the discussion.
4. The maxim of manners - when one tries to be as clear, brief, and orderly as one can in what one-says, and where one avoids obscurity and ambiguity.

As Pitik underwent iterative testing, it was progressively trained to comprehend the optimal and most tolerable responses gathered from users during the Data Collection phase. Currently, Pitik operates in alignment with the principles of the Cooperative Principle, particularly Grice's Maxims, while also drawing on the insights from our collected data. Violations of Grice's Maxims are triggered whenever DialogFlow fails to match a user input with a defined intent, resulting in a fallback intent. This phenomenon predominantly affects Pitik's generic questions regarding exercise, diet, smoking, and alcohol, which often elicit a wide range of user responses, frequently resulting in the flouting of the Cooperative Principle. As DialogFlow matches the closest training phrases, it struggles to address variations in user responses. To enhance functionality, developers utilize intents to define specific tasks that users can perform, ensuring a more structured interaction.

Figure 2. *Example of Grice Maxim Occurrences in Pitik*

Process Flow of Pitik using DialogFlow

The researchers utilized DialogFlow, a natural language understanding platform, to design and integrate a conversational user interface for the Pitik chatbot. The platform facilitates user interaction by allowing end-users to input text, which DialogFlow matches to specific intents while extracting relevant parameters. Once the intent is identified, DialogFlow sends a webhook request to the designated service, including information about the matched intent, the associated action, parameters, and the predefined response. Pitik then executes the necessary actions, retaining the extracted information and guiding users through the relevant conversational flow. Subsequently, Pitik generates a webhook response message directed back to DialogFlow, containing the response intended for the end-user. Finally, DialogFlow relays this response to the user, ensuring seamless interaction and effective communication throughout the conversation.

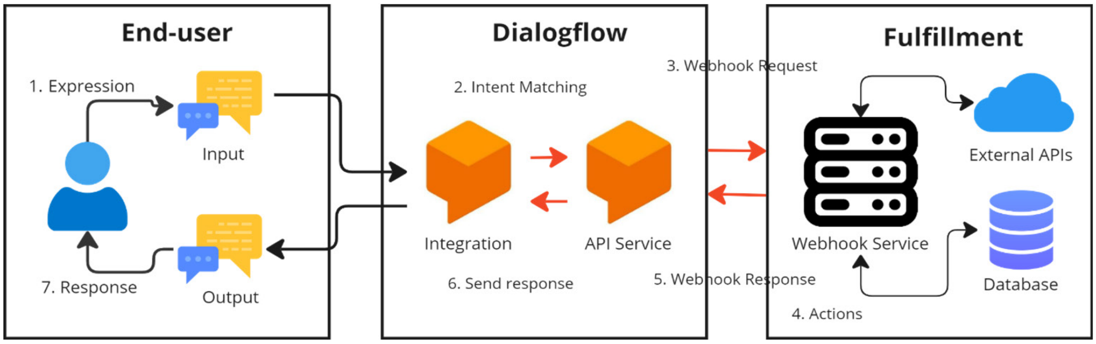

Figure 3. *Process Flow of Pitik*

Cardiovascular Risk Assessment Sheet

Pitik initiates the conversation by greeting the user, establishing a friendly and welcoming environment. Following the exchange of greetings, Pitik collects essential demographic information, including the user's name, age, sex, weight, and height. For female users, the chatbot then inquires about pregnancy-related details, such as current pregnancy status and the number of children they have. After gathering this information, Pitik collects information about the user's symptoms, and medical, surgical, and family histories. Subsequently, the chatbot poses questions related to laboratory results, including measurements such as HbA1c, systolic blood pressure, and diastolic blood pressure. During the contact details stage, users can provide their email addresses to receive comments or recommendations from healthcare professionals. Finally, Pitik generates a risk factor assessment and calculates the user's Body Mass Index (BMI). Pitik's complete conversational flow is illustrated in Figure 4.

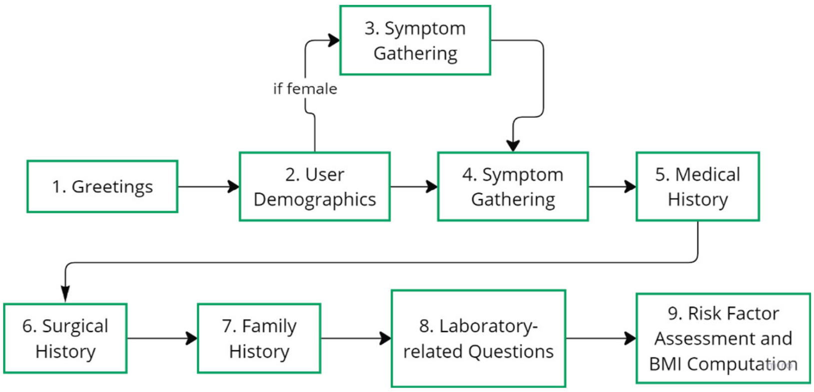

Figure 4. *Pitik Conversation Flow*

To assess the cardiovascular risk of Pitik users, the researchers utilized a cardiovascular risk assessment sheet[1] recommended by a healthcare professional conducting community-based profiling and risk factor assessment. Risk factor is then computed using the Framingham formula (D'Agostino et al., 2008):

Risk Factors = (ln(Age) * 3.06117) + (ln(Total_cholesterol) * 1.12370) - (ln(HDL_cholesterol) * 0.93263) + (ln(Systolic_blood_pressure) * On_blood_pressure_medication) + Cigarette_smoker + Diabetes_present - 23.9802

$$\text{Risk} = 100 * (1 - 0.88936^{e(\text{Risk\_Factors})})$$

Pitik is trained to recognize input symptoms based on a predefined list of cardiovascular-related diseases and their corresponding symptoms[2]. Table 1 shows the cardiovascular diseases and their corresponding symptoms (in Cebuano-Binisaya).

Table 1. *Cardiovascular Diseases and Symptoms*

Cardiovascular Disease	Symptoms
Congenital Heart Disease	*Pagkablue, nanluspad, hubag, hupong,lisud ginhawa, wad-an kusog, pitik, di regular, kasingkasing, lipong, sakit ulo.*
Arrythmia	*Gaan - Paminaw, Bloat - Tiyan, Paspas - Heartbeat, Hinay - Pulso, Taas - Blood presssure, Diabetes, Sakit - dughan, Maluya, Kalipong, Kapit-os ug kabalaka, Alcohol - caffeine*
Cardiomyopathy	*kapoy, lisud og ginhawa, hubag ang bitiis, bukong ug tiil, bloat nga tiyan, ubo, sakit sa dughan, lipong, luya*
Atherosclerosis	*kabugnaw,manhid,Kahuyang,kawad-an og usog*
Coronary Heart Disease	*Sakit sa dughan, Dili komportable ang dughan, Kumot sa dughan, Kakulang sa ginhawa, Kasukaon*
Heart Infection	*Kakapoy, Kalipong, Pagkulbakulba Sakit sa dughan, Kalintura, Nagamaga akong bitiis, Dugo sa akong ihi, pula na tuldok sa lawas, sa puti sa mata o sa sulod sa baba nay pula pula.*

Figure 5 shows an actual example of how Pitik attempts to capture user symptoms.

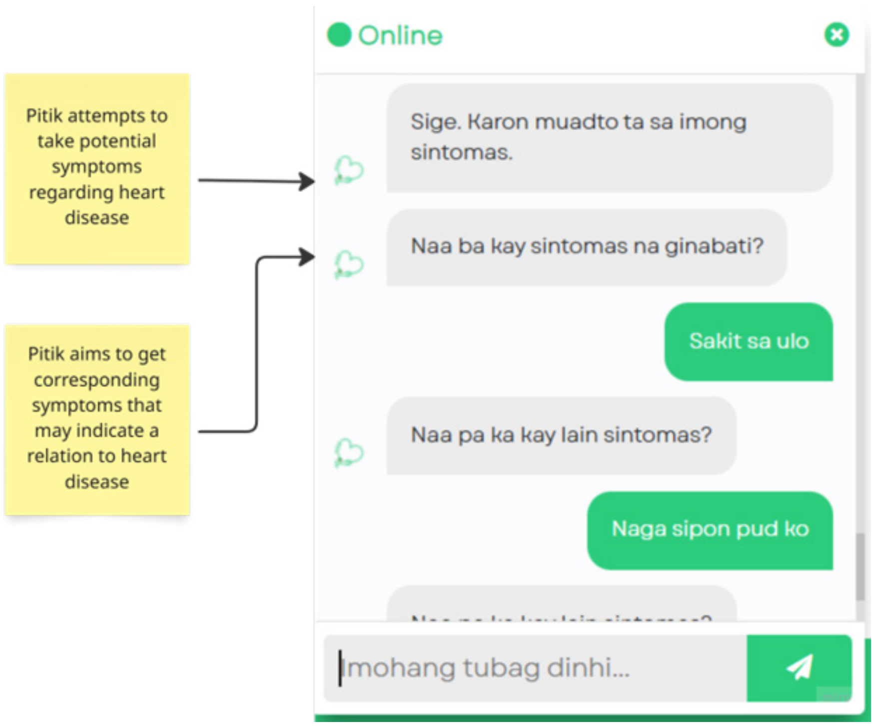

Figure 5. *Pitik Gets User Symptoms*

Analytic Hierarchy Process

Pitik's testing phase centered on evaluating the chatbot's overall quality using the method recommended by Radziwill and Benton (2017). This approach employs the Analytic Hierarchy Process (AHP), a structured technique for organizing and analyzing complex decisions involving qualitative and quantitative factors. The researchers selected AHP to assess and quantify criteria related to the quality of the Pitik chatbot effectively. The goal was to compare and evaluate the performance of the original version of Pitik against the updated version per iteration. The criteria chosen for this assessment were aligned with the study's thematic focus and objectives shown in Figure 6.

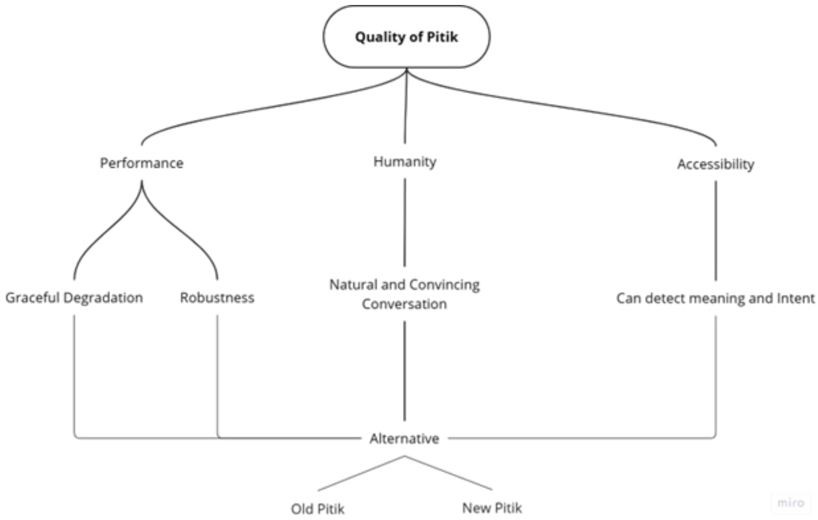

Figure 6. *Hierarchical Structure for Pitik Chatbot Evaluation*

Each category has its own evaluation methods to contribute to the overall chatbot quality. For the category 'Humanness', the evaluation made use of a Google form comprising an answerable, open-ended question form and a Likert-scale format question appearing at the very end of the chat conversation that lets users be able to evaluate the chatbot based on the category that is identified to analyze the quality producing Likert scale information and opinionated information about the user-opinion. The researchers utilized a tally system to count Gricean Maxim Violations for the user inputs and a linguist expert-reviewed verdict. For the 'Performance', the researchers used another tally system that counts the number of times Pitik has failed to respond appropriately during the conversation with the user. For the 'Accessibility', the researchers used a similar method to find the Performance with the difference of tallying the circumstances when the users are the ones to ask the questions for Pitik.

We have used Precision, Recall, F1 Score, Accuracy, and its weighted average to evaluate our post-intent matching algorithm which trained several machine learning models to classify intents. Precision, the ratio between the True Positives and all the Positives, enables us to identify true positives concerning the overall positive predictions. A high precision means it has a lower false positive rate. Recall is the ratio of the correctly predicted positive occurrences with respect to the overall observations in the actual class. F1 Score is the average from both Precision and Recall and in return, considers false positives and negatives. Accuracy is the number of correct predictions concerning the overall data. This is achieved by adding true positives and negatives to true positives, true negatives, false negatives, and false positives. The weighted average is generated by multiplying each evaluation measurement by its corresponding weight which is the percentage count of the total class concerning the overall test data.

RESULTS

The researchers built a schema-guided dialogue set gathered from 100 participants by reaching out through social media and word of mouth. The selection of participants preferred to be those residing in provincial areas, but due to limited time and accessibility, the selection was made based on whether the participant could communicate using the identified language of Pitik. The first 50 participants interacted with Pitik during the second iteration, and then the remaining 50 participants interacted with Pitik during the third iteration. A link was sent to the participants with instructions and information concerning the purpose of the study.

The researchers gathered the responses and could group and analyze the data by checking which responses violated Gricean Maxims with the help of an expert. An example is shown in Table 2, which produced bugs in the system and how the participants were able to answer specifically the laboratory-related questions such as blood pressure.

Table 2. *Gricean Maxims Violations*

Gricean Maxims	Sample Responses	2nd Iteration Violations	3rd Iteration Violations
Manner	*"murag naa sa 125-250", "sauna naa pero karon wala"*	44	22
Relation	*"wala kay ginabutang ra nko sa kamot", "gipaak kog ilaga, dugay na sukad bata pa ko"*	18	6
Quantity	*"oo pag ma stress ko", "dli kay weak"*	24	16
Quality	*"murag naa sa 125-250", "5'2 ata"*	7	5

Several models were trained to develop an effective classifier for intent recognition. The researchers evaluated and compared the performance of four algorithms: Naive Bayes (NB), Support Vector Machine (SVM), Multi-Layered Perceptron (MLP), and Recurrent Neural Network (RNN). These algorithms are well-established for classifying text data. Naive Bayes classifiers operate under the assumption of strong (or naive) independence between the attributes of data points. They are widely used such as spam filtering, text analysis, and medical diagnosis applications. Support Vector Machines (SVM) are supervised machine learning models designed for binary classification tasks, separating data into two groups. The Multi-Layered Perceptron (MLP) classifier relies on an underlying neural network to perform classification, allowing it to learn complex patterns in the data. On the other hand, Recurrent Neural Networks (RNNs) are commonly employed in speech recognition and natural language processing due to their ability to recognize sequential patterns in data and predict future scenarios. However, they are typically more complex to implement.

Table 3 compares the four classifiers across two domains: Diet/Exercise and Smoking/Alcohol. All models were tested on a balanced dataset comprising 40 records from 200 training samples.

Table 3. *Model Performances for Sample Diet/Exercise and Smoking/Alcohol Areas*

Area of Comparison	Models	Precision	Recall	F1 Score	Accuracy
Diet/Exercise	NB	71	70	**70**	70
	SVM	70	70	**70**	70
	MLP	65	65	65	65
	RNN	58	57	57	57
Smoking/Alcohol	NB	72	68	66	68
	SVM	78	72	**71**	73
	MLP	70	68	66	68
	RNN	61	60	59	60

The researchers assessed accessibility by examining the success rate of five critical laboratory questions, such as HBA1C and Systolic Blood Pressure, which often pose significant barriers due to the need for prior medical knowledge and specific instrument results. Initially, the average success rate was a mere 22%, as most respondents lacked lab results, leading to widespread failure in answering these questions. However, in the third iteration, the introduction of the Naive Bayes algorithm and post-intent processing significantly boosted both performance and accessibility, with rates soaring from 65% to 79% and 22% to an impressive 96%, respectively. This remarkable improvement was driven by the third iteration's enhanced user guidance and response processing, featuring new tools like suggestion chips and tooltips. The researchers recognized that many users sought definitions for medical terminologies, such as HBA1C, prompting the integration of tooltips to assist users unfamiliar with terms like Systolic Blood Pressure, Diastolic Blood Pressure, Total Cholesterol, and HDL Cholesterol. Additionally, suggestion chips were implemented to address instances where the chatbot failed to recognize user input, guiding users toward a format the system could understand. For example, when a user responded "n/a" to a question about waist circumference, a suggestion chip appeared to offer guidance. These enhancements were crucial in overcoming the limitations encountered during the training phase, ensuring that the system could adapt to unforeseen inputs and vastly improving overall user experience.

There are two key scenarios where suggestion chips are strategically deployed to enhance user interaction. In the first scenario, suggestion chips appear during the initial occurrence of a question to confirm the accuracy of user input. These questions typically involve a prompt to verify that the chatbot has correctly understood the input. In the second scenario, suggestion chips address questions that may elicit unexpected or unanticipated responses, which became evident during the system's second iteration. For instance, when Pitik asked for the father’s name, one user out of 100 responded that they did not know their father. The researchers created an Intent that acknowledges and skips the question to accommodate such rare but important cases. However, due to the limited occurrence of such responses, there weren’t enough training phrases to fully support this Intent. Therefore, suggestion chips were implemented at the first occurrence of such questions, enabling users to skip questions about parents they do not know. This adjustment significantly improved performance by allowing the chatbot to handle a broader range of responses while maintaining focus on essential information.

Moreover, many did not know their results when users encountered one of the five laboratory questions—HBA1C, Systolic and Diastolic Blood Pressure, Total Cholesterol, and HDL Cholesterol. Instead of persisting with these questions, the researchers provided suggestion chips, allowing users to select which laboratory questions they could answer. These enhancements improved user experience and significantly increased accessibility as the chatbot became more adept at catering to diverse user needs and circumstances.

The data collected from the respondents were analyzed using the Analytic Hierarchy Process (AHP) to assess potential improvements in version performance. Performance metrics were derived from instances where Pitik triggered a fallback intent and failed to progress toward subsequent intents. The key quality attribute under performance is robustness, particularly its ability to handle unexpected inputs and exhibit graceful degradation. This includes its capacity to resume conversations seamlessly and allow users to skip questions they cannot answer. The second evaluation criterion, humanness, was assessed based on respondents’ feedback and their subjective ratings of Pitik’s conversational experience. The primary quality attribute within this category is its ability to maintain a coherent and contextually relevant discussion. Finally, the accessibility category was evaluated using a similar approach to performance assessment, focusing specifically on user responses related to accessibility aspects. This ensured a comprehensive analysis of Pitik's usability and inclusivity for diverse user needs.

DISCUSSION

The Naive Bayes and SVM models demonstrate similar performance in the Diet/Exercise category, each achieving Precision, Recall, F1 Score, and Accuracy values of approximately 70%. This suggests that these models are relatively effective in classifying or predicting outcomes related to Diet and Exercise. The MLP model shows a slight decline in performance, with all metrics standing at 65%, indicating a moderately lower ability to generalize in this context. Conversely, the RNN model exhibits the weakest performance, with metrics in the of 57-58% range, highlighting significant challenges in capturing the patterns associated with

Diet and Exercise. In the Smoking/Alcohol category, the SVM model emerges as the most effective, boasting the highest Accuracy (73%) and relatively high values across all other metrics, underscoring its robustness in handling this classification task. The Naive Bayes model, while strong in Precision (72%), shows a lower F1 Score (66%), indicating some trade-offs between Precision and Recall. MLP's performance mirrors that of Naive Bayes in this context, while the RNN model continues to underperform, with all metrics falling around 59-61%. The results indicate that the SVM model is particularly well-suited for tasks related to Smoking/ Alcohol, outperforming other models by a notable margin. Its ability to maintain high Accuracy and balanced metrics suggests that it can be relied upon for more accurate predictions and classifications in this domain. On the other hand, the RNN model's consistent underperformance across both areas suggests that it may struggle with the type of data or features used in these tasks.

Table 4. *Reciprocal Matrix for Pairwise Comparisons*

Category	Performance	Humanness	Accessibility
Performance	1	9	7
Humanness	0.111	1	0.5
Accessibility	0.143	2	1
Criterion Weight	0.790	0.077	0.133

Table 4 highlights how each category is weighted relative to the others. Performance is deemed significantly more important than both Humanness and Accessibility, with a comparison ratio of 9:1 against Humanness and 7:1 against Accessibility. Conversely, Humanness is considered less important than Accessibility, with a ratio of 0.5:1, indicating that Accessibility is valued twice as much as Humanness. The reciprocal nature of the matrix is evident, as each off-diagonal value corresponds to the reciprocal of its counterpart. For instance, the value comparing Humanness to Performance (0.111) is the reciprocal value comparing Performance to Humanness (9). The final row of the table reveals the derived criterion weights, reflecting the overall importance of each category: Performance carries the highest weight at 0.790, indicating its dominant role in the decision-making process. Accessibility follows with a weight of 0.133, while Humanness has the lowest weight at 0.077. These weights suggest that Performance is overwhelmingly prioritized, with Accessibility also considered important. However, Humanness is emphasized less, showing a strong preference for functional aspects over human-centric qualities.

CONCLUSION

This study developed Pitik, a Cebuano-Binisaya intent-based chatbot designed for cardiovascular disease patient profiling and risk factor assessment, integrating technical advancements in intent detection and response generation through the Diag-Ex framework. A key contribution of this research is the creation of a schema-guided dialogue dataset specific to cardiovascular health in Cebuano-Binisaya, addressing a critical gap in digital healthcare resources for underserved communities. The Diag-Ex framework, composed of Pre-Intent Matching and Post-Intent Matching algorithms, significantly improved Pitik's intent detection accuracy and conversational efficiency. Pre-Intent Matching, utilizing regular expressions, enhanced Pitik's ability to validate user input before intent processing, minimizing mismatches and improving data consistency. Meanwhile, Post-Intent Matching, initially implemented with Multinomial Naive Bayes, was further optimized by adopting Support Vector Machine (SVM), which demonstrated superior classification performance for small datasets. The integration of suggestion chips and tooltips further enhanced user interaction and accessibility, leading to a notable reduction in Gricean Maxim violations and improved response accuracy. Evaluation results showed that these advancements led to substantial improvements in performance (from 64.78% to 79.44%) and accessibility (from 22.28% to 95.83%) between the second and third iterations. The structured application of the Analytic Hierarchy Process (AHP) provided quantitative validation of these improvements, confirming that performance, accessibility, and humanness were effectively balanced.

Recommendations

To further optimize Pitik, several enhancements are recommended. First, algorithm optimization is necessary, as the study found that SVM outperformed Multinomial Naive Bayes in post-intent matching. Future iterations should exclusively use SVM or explore deep learning models to achieve even greater accuracy in intent classification. Second, improving conversational flow is crucial, as users report fatigue due to lengthy interactions. Implementing a button-based response system can streamline engagement and enhance user experience. Third, enhanced medical support features should be integrated. While Pitik effectively provides risk factor assessments, users expressed the need for personalized consultations and medical guidance. To address this matter, future versions should include doctor-editable risk assessment forms and deliver evidence-based recommendations based on user inputs. Finally, dataset expansion is essential, as the scarcity of Cebuano-Binisaya medical datasets limits the chatbot's adaptability. Continuous efforts should be made to broaden and refine the training data, improving Pitik's responsiveness to diverse patient queries. By addressing these recommendations, Pitik can continue to advance as a technologically robust and linguistically accessible digital healthcare tool, effectively bridging the gap in cardiovascular disease profiling for Cebuano-Binisaya speakers.

Author Contributions

Cedeño: Software. Manteza: Resources, Investigation; **Nacar**: Data Curation, Writing (Original Draft); **Umbukan:** Software, Formal Analysis; **Muaña:** Data Curation, Domain Expert (Pharmacy); **Cruz:** Validation, Data Curation; **Benablo:** Writing (Review and Editing); **Adlaon:** Supervision, Project Administration, Writing (Review and Editing)

Funding

This research received no external funding.

Ethical Approval

Not Applicable

Competing interest

The authors declare no conflicts of interest.

Data Availability

Data will be made available by the corresponding author upon request.

Declaration of Artificial Intelligence Use

In this work, the author(s) utilized artificial intelligence (AI) tools and methodologies, ChatGPT, to refine paragraph statements. After using this tool/service, the author(s) evaluated and revised the content as necessary and take(s) full responsibility for the published content.

REFERENCES

Cacciata, M. C., Alvarado, I., Jose, M. M., & Evangelista, L. S. (2021). Health determinants and risk factors for coronary artery disease among older Filipinos in rural communities. European Journal of Cardiovascular Nursing, 20(6), 565-571. https://doi.org/10.1093/eurjcn/zvaa039

Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016). Using electronic health records for population health research: A review of methods and applications. Annual Review of Public Health, 37, 61-81. https://doi.org/10.1146/annurev-publhealth-032315-021353

D'Agostino, R. B., Sr., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: The Framingham Heart Study. Circulation, 117(6), 743-753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579 .

Evans, R. S. (2016). Electronic health records: Then, now, and in the future. Yearbook of Medical Informatics, 25(S01), S48-S61. https://doi.org/10.15265/IYS-2016-s006.

Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics, 3, 41-58. Academic Press.

Iyanna, I. S., Kaur, K. P., Ractham, P., Talwar, S., & Islam, A. K. M. N. (2022). Digital transformation of the healthcare sector: What is impeding adoption and continued usage of technology-driven innovations by end-users? Journal of Business

Research, 153, 150-161. https://doi.org/10.1016/j.jbusres.2022.08.007
Liwanag, L. M. A., Liwanag, G. L., & Liwanag, L. A. (2024). AI in Anthem: A comparative analysis of the English and Filipino ChatGPT 4 translations from the existing translations of the Philippine National Anthem. Recoletos Multidisciplinary Research Journal, 12(2), 91-102. https://doi.org/10.32871/rmrj2412.02.07
Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. Risk Management and Healthcare Policy, 4, 47-55. https://doi.org/10.2147/RMHP.S12985
Noceda, A. V. G., Acierto, L. M. M., Bertiz, M. C. C., Dionisio, D. E. H., Laurito, C. B. L., Sanchez, G. A. T., & Loreche, A. M. (2023). Patient satisfaction with telemedicine in the Philippines during the COVID-19 pandemic: A mixed methods study. BMC Health Services Research, 23(1), 277. https://doi.org/10.1186/s12913-023-09127-x .
Quinn, M., Forman, J., Harrod, M., Winter, S., Fowler, K. E., Krein, S. L., Gupta, A., Saint, S., Singh, H., & Chopra, V. (2019). Electronic health records, communication, and data sharing: Challenges and opportunities for improving the diagnostic process. Diagnosis (Berlin), 6(3), 241-248. https://doi.org/10.1515/dx-2018-0036
Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv:1704.04579. https://doi.org/10.48550/arXiv.1704.04579 .
Reyes, A. T., Serafica, R., Kawi, J., Fudolig, M., Sy, F., Leyva, E. W. A., & Evangelista, L. S. (2023). Using the socioecological model to explore barriers to health care provision in underserved communities in the Philippines: A qualitative study. Asian/Pacific Island Nursing Journal, 7, e45669. https://doi.org/10.2196/45669
Sarma, M., Chatterjee, S., Mohanty, S., Puravankara, R., & Bali, M. (2019). Diabot: A predictive medical chatbot using ensemble learning. International Journal of Recent Technology and Engineering, 8(2), Article B2196. https://doi.org/10.35940/ijrte.B2196.078219
Tan, R., Kawaja, A., Ooi, S. P., & Ng, C. J. (2024). Communication barriers faced by pharmacists when managing patients with hypertension in a primary care team: a qualitative study. BMC Primary Care, 25(1). https://doi.org/10.1186/s12875-024-02349-w

How to cite this article:
Cedeño, J., Manteza, A., Nacar, N., Umbukan, M., Muaña C., Vasay-Cruz, M. J., Benablo, C. I., & Adlaon, K. M. (2025). Pitik: A Cebuano-Binisaya Intent-Based Chatbot for Cardiovascular Disease Patient Profiling and Risk Factor Recommendations. Recoletos Multidisciplinary Research Journal 1(Special Issue), 63-75. https://doi.org/10.32871/rmrj25si.t2407