

Development and Evaluation of An Explainable AI-Based Chatbot for First-Aid Assistance in Domestic Accident Scenarios among Ebonyi State Indigenes

Jeremiah, Chukwu¹, Agwu Joy Nneka², Nweke Chukwuma², Nwode, Sunday Nnamdi²

¹Department of Computer Science, Ebonyi State University, Abakaliki – Nigeria

²ICT/Research Unit, Ebonyi State University, Abakaliki – Nigeria

ABSTRACT

Domestic accidents constitute an important source of preventable injury and may require immediate first-aid intervention before professional medical care is available. Artificial intelligence (AI)-enabled conversational systems offer a potential mechanism for delivering accessible first-aid guidance; however, their reliability, transparency, and ability to correctly identify accident scenarios remain important concerns. This study developed and evaluated an explainable AI-based chatbot for first-aid assistance in domestic accident scenarios, with the long-term goal of supporting accessible first-aid decision support among populations in Ebonyi State, Nigeria. A natural language processing and machine-learning framework was developed for first-aid intent classification. A labelled first-aid dataset comprising 93 user utterances across 21 injury and emergency-related classes was used for the preliminary experiment. Text was normalized and represented using Term Frequency–Inverse Document Frequency (TF-IDF) features with unigram and bigram representations. Four classifiers—Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Multinomial Naïve Bayes (NB)—were evaluated using four-fold stratified cross-validation. Performance was assessed using accuracy, precision, recall, and F1-score, while a separate stratified hold-out set was used to examine the generalization performance of the best-performing classifier. Explainability and emergency-escalation mechanisms were incorporated into the proposed chatbot architecture. SVM produced the strongest overall performance, achieving 85.96% accuracy, 80.95% precision, 86.31% recall, and 82.10% F1-score. LR achieved 66.71% accuracy, 61.31% precision, 66.07% recall, and 61.87% F1-score; RF achieved 59.19%, 52.29%, 59.52%, and 53.04%, respectively; while NB achieved 36.55%, 28.26%, 36.31%, and 29.95%. SVM achieved 83.33% accuracy on the independent hold-out evaluation. The findings demonstrate the feasibility of machine-learning-based first-aid intent classification as a foundation for an explainable conversational first-aid support system. SVM demonstrated superior performance among the evaluated models. Nevertheless, the preliminary dataset size and class distribution limit generalizability. Future research should incorporate larger, expert-annotated and culturally representative Nigerian datasets, transformer-based models, clinical expert validation, explanation-quality assessment, and prospective user-safety evaluation before real-world deployment.

Keywords: Explainable Artificial Intelligence, chatbot, first aid, domestic accidents, emergency care, natural language processing

Date of Submission: 16-09-2026

Date of Acceptance: 30-09-2026

I. Introduction

Domestic accidents are unexpected events occurring within or around the household that may result in injury, illness, disability, or death. Common examples include burns, cuts, falls, fractures, choking, poisoning, electrical shock, drowning, animal bites, and injuries resulting from household equipment. Although many of these incidents can be prevented, immediate intervention remains essential when an accident occurs. First aid provides the initial response while the affected individual awaits definitive medical care.

The American Heart Association defines first aid as helping behaviour and initial care provided for an acute illness or injury, emphasizing recognition, assessment, prioritization, appropriate intervention, recognition of limitations, and escalation for additional care when necessary (AHA, 2024). Similarly, international first-aid guidance emphasizes preservation of life, alleviation of suffering, prevention of further injury, and promotion of recovery (Zideman *et al.*, 2020). This makes first aid particularly important in domestic environments where professional healthcare personnel may not be immediately available.

The problem is particularly relevant in low- and middle-income settings, where delays in emergency response and limitations in prehospital services may increase the consequences of injuries. Research conducted in Nigeria has demonstrated deficiencies in first-aid knowledge among members of the public. According to Eze *et al.* (2012) in the study of commercial drivers in Nigeria found that only 37.6% correctly identified the appropriate sequence for basic resuscitation priorities, although 79.9% indicated willingness to participate in first-aid training. Similarly, first-aid education has been shown to improve first-aid skills among Nigerian commercial drivers, demonstrating the potential value of accessible interventions for strengthening layperson emergency response (Balogun *et al.*, 2015).

The domestic environment presents an additional challenge. Domestic accidents can produce serious injuries, particularly among children. Nigerian evidence has documented household injuries involving falls and traumatic brain injuries among children, highlighting the importance of prevention and appropriate responses to domestic accidents (Buowari & Ikpa, 2022). Such evidence supports the need for accessible first-aid information that can be obtained rapidly at the point of need.

Recent developments in artificial intelligence provide an opportunity to address this challenge. Healthcare chatbots can interact with users using natural language and provide information without requiring users to navigate complex healthcare information systems. A systematic review of healthcare chatbots reported their growing application for health information, support, and communication, particularly where accessibility and user interaction are important considerations (Grassini *et al.*, 2025).

However, the use of AI chatbots for emergency and first-aid guidance introduces significant safety concerns. Langdorf *et al.* (2024), in an experimental comparison of four large language model chatbots responding to emergency-care questions, identified inaccuracies and potentially dangerous information in chatbot responses. Their findings demonstrate that the ability of an AI system to produce fluent and plausible language does not necessarily guarantee clinically safe recommendations. Similar concerns have been identified in research evaluating large language models for health advice, where substantial variation exists in reporting quality, clinical evaluation, and attention to patient-safety considerations (Huo *et al.*, 2025).

Consequently, an AI system designed for first-aid assistance should not function as an unrestricted generative chatbot. It should incorporate a controlled medical knowledge base, explicit safety rules, uncertainty management, escalation mechanisms, and explanations for its recommendations.

Explainable Artificial Intelligence (XAI) provides an important mechanism for addressing this requirement. XAI seeks to make AI decisions understandable to users by providing information about how and why a particular output was generated. In healthcare, explainability is particularly important because opaque AI systems may reduce trust and make it difficult to identify errors or inappropriate recommendations (Eke & Shuib, 2025). Reviews of XAI-based clinical decision-support systems similarly emphasize the importance of interpretability, transparency, and human-centred explanations in healthcare applications (Hong *et al.*, 2024).

Against this background, this study proposes an Explainable AI-Based Chatbot for First-Aid Assistance in Domestic Accident Scenarios among Ebonyi State Indigenes. The proposed system is intended to provide immediate, understandable, evidence-based first-aid guidance while recognizing situations that require urgent professional medical intervention.

II. Problem Statement

Domestic accidents can occur without warning, and the first few minutes following an injury may be critical. In many households, however, the person responding to an accident may have limited first-aid knowledge. Incorrect interventions, delays, inappropriate home remedies, or failure to recognize a life-threatening condition may worsen the victim's condition.

The problem is compounded by the accessibility limitations of conventional healthcare information. A person experiencing an emergency may need to search multiple websites or consult different sources before determining what to do. This process can be inappropriate during time-sensitive emergencies.

Although AI chatbots provide a potentially convenient interface for accessing health information, existing evidence indicates that unrestricted AI-generated medical advice may contain omissions, inaccuracies, and dangerous recommendations. Langdorf *et al.* (2024) found potentially dangerous information in a proportion of chatbot responses to emergency-care questions. Birkun and Gautam (2024) similarly reported that large language model-based chatbot responses to first-aid questions concerning heart attacks could contain omissions, inaccuracies, and misleading instructions.

Another important problem is lack of explainability. A chatbot may tell a user what to do without explaining why the recommendation is appropriate or when it should not be followed. In safety-critical healthcare applications, such opacity may undermine user trust and make it difficult to detect inappropriate recommendations.

Furthermore, existing AI healthcare chatbot research is largely general-purpose and does not necessarily address the cultural, linguistic, technological, and healthcare realities of specific Nigerian populations. There is

therefore a need for a context-sensitive system that focuses specifically on domestic accidents and incorporates explainability, safety constraints, and appropriate referral mechanisms.

III. Literature Review

2.1 Concept of First Aid and Domestic Accidents

First aid represents the immediate assistance provided to an individual who has sustained an injury or developed an acute illness before definitive medical treatment becomes available. The American Heart Association (AHA) and American Red Cross define first aid as helping behaviours and initial care provided for acute illness or injury. The scope of first aid includes recognizing and assessing the condition, prioritizing the need for intervention, providing appropriate care within the responder's competence, recognizing limitations, and seeking additional medical assistance when necessary (American Heart Association [AHA], 2024).

First aid is particularly important because the interval between injury occurrence and professional medical intervention may determine the severity of the outcome. The WHO-ICRC Basic Emergency Care framework emphasizes systematic initial assessment and management of time-sensitive conditions, particularly in resource-limited settings where early intervention may substantially influence outcomes (World Health Organization [WHO] & International Committee of the Red Cross [ICRC], 2018).

Domestic accidents are injuries or acute health events occurring within the home or its immediate environment. They may result from falls, burns, cuts, choking, poisoning, electrical shocks, drowning, animal bites, and other household hazards. Although some domestic injuries are minor, others can be life-threatening and require immediate intervention. The 2024 AHA/Red Cross guidelines specifically recognize residential settings as important contexts for first-aid intervention and cover traumatic, medical, environmental, and toxicological emergencies (AHA, 2024).

The domestic setting is particularly important for first-aid interventions because professional assistance may not be immediately available. A household member, neighbour, or other bystander is frequently the first person to respond. Consequently, the effectiveness of first aid depends partly on whether the responder can recognize the emergency, understand what action should be taken, avoid harmful interventions, and determine when professional assistance is necessary.

The contemporary first-aid approach therefore extends beyond simply knowing individual procedures. It requires the ability to recognize, assess, prioritize, intervene, and escalate. This principle has direct implications for an AI-based first-aid chatbot because the system should not simply provide an answer to a user's question; it should support the sequence of recognition, assessment, appropriate intervention, and referral.

2.2 First-Aid Knowledge and the Need for Accessible Decision Support

The effectiveness of first aid depends substantially on the knowledge and confidence of potential first responders. In real-world emergencies, individuals may experience anxiety, uncertainty, and time pressure, all of which can interfere with appropriate decision-making. A person who has never received formal first-aid training may know that an injury requires attention but may not know the appropriate intervention or may rely on traditional or potentially harmful remedies.

The 2024 AHA/Red Cross guidelines emphasize that first-aid providers must operate within their knowledge and skill limitations and seek additional care when necessary (AHA, 2024). This principle suggests that an effective digital first-aid system should not encourage users to undertake interventions beyond their competence. Instead, it should provide simple, sequential instructions and clearly identify circumstances in which professional medical attention is required.

This requirement is particularly relevant to AI-enabled first-aid assistance. An intelligent system could potentially reduce the information-access barrier by allowing users to describe an accident using ordinary language rather than requiring them to understand medical terminology. For example, a user may type "*my child poured hot water on his hand*" rather than identifying the event as a scald injury. An NLP-based system could interpret the statement, identify the likely accident category, ask relevant follow-up questions, and provide appropriate first-aid guidance.

The proposed chatbot therefore extends the conventional concept of digital health information retrieval toward interactive decision support. Rather than presenting static information about burns, bleeding, falls, or poisoning, it can potentially adapt its response according to the user's description, symptoms, severity indicators, and available information.

2.3 Artificial Intelligence in Healthcare

Artificial intelligence has increasingly been applied to healthcare tasks involving prediction, diagnosis, clinical decision support, patient education, information retrieval, and communication. Machine-learning systems can identify patterns in large datasets, while natural language processing enables computational systems to interpret and generate human language.

The increasing sophistication of generative AI and large language models has expanded the potential role of conversational systems in healthcare. Unlike conventional rule-based systems, large language models can interpret diverse forms of natural-language input and generate responses that resemble human conversation. This capability makes conversational AI potentially useful for patient education and health information services.

However, healthcare represents a particularly sensitive domain for AI deployment because errors can result in physical or psychological harm. Consequently, AI performance cannot be evaluated solely according to linguistic fluency or conventional classification accuracy. Healthcare AI requires consideration of clinical correctness, safety, transparency, reliability, privacy, fairness, and human oversight.

This distinction is particularly important for first-aid applications. A chatbot that produces a fluent but clinically inappropriate recommendation could potentially cause a user to delay emergency treatment or undertake an unsafe intervention. Thus, the proposed first-aid chatbot requires a more controlled architecture than a general-purpose conversational AI system.

2.4 Healthcare Chatbots

Healthcare chatbots are conversational computer systems designed to communicate with users through natural language and provide health-related information, guidance, education, or support. Their potential advantages include continuous availability, conversational interaction, scalability, and the ability to provide information without requiring immediate contact with a healthcare professional.

The development of conversational AI has created new possibilities for health information delivery because users can communicate with systems using ordinary language. This may be particularly useful for people who find conventional medical websites difficult to navigate or who require information rapidly.

However, the evidence also indicates that healthcare chatbots require rigorous evaluation. In emergency-care contexts, a system must distinguish between conditions that can receive basic first aid and conditions requiring immediate professional intervention. This creates a substantially higher safety requirement than that associated with general health education.

Langdorf et al. (2024) evaluated responses from four large language model chatbots to questions involving emergency-care conditions. Their study identified important limitations in AI-generated emergency-care advice. A subsequent analysis reported that potentially dangerous information occurred in approximately 5%–35% of responses, while information concerning when to seek urgent or emergency care was frequently incomplete or inaccurate.

These findings are particularly relevant to the proposed study. They suggest that a first-aid chatbot should not simply be an unrestricted large language model. Instead, the system should incorporate controlled medical knowledge, safety rules, emergency triggers, uncertainty handling, and expert validation.

2.5 Artificial Intelligence for First-Aid Assistance

The application of AI to first-aid assistance is an emerging area at the intersection of healthcare informatics, conversational AI, and emergency decision support. The potential value of such systems arises from their ability to provide information immediately and interactively at the point where an accident occurs.

A conventional first-aid application may provide a predefined list of procedures. An AI-based system can potentially improve interaction by asking questions based on the user's responses. For instance, after a user reports a burn, the system could ask about the cause, affected body part, size of the affected area, presence of blistering, and other relevant indicators before providing guidance.

This interactive approach is important because first-aid decisions are often context-dependent. The appropriate response to a minor superficial burn may differ substantially from the response to an extensive burn, an electrical burn, or a burn involving a sensitive anatomical area.

Nevertheless, the use of AI for first aid raises a critical question: How can an AI system provide useful assistance without generating unsafe medical advice?

This question represents an important research problem. The evidence from emergency-care chatbot evaluations indicates that generative AI can produce plausible but unsafe recommendations (Langdorf et al., 2024). Therefore, the proposed system will combine AI-based language understanding with a validated first-aid knowledge base and rule-based safety layer.

This hybrid architecture can retain the conversational advantages of AI while restricting recommendations to approved interventions.

2.6 Explainable Artificial Intelligence

Explainable Artificial Intelligence refers to approaches that make the behaviour, predictions, or decisions of AI systems understandable to human users. Explainability has become increasingly important in healthcare because many high-performing AI models operate as black boxes whose internal decision processes are difficult for users to understand.

Eke and Shuib (2025) argue that opacity, bias, robustness, and generalizability remain important barriers to trust in healthcare AI and that transparency and explainability are important components of trustworthy AI adoption.

A systematic review by Aziz *et al.* (2024) examined XAI applications in clinical decision-support systems and identified continuing challenges involving datasets, evaluation approaches, interpretability methods, and interdisciplinary collaboration. More recent meta-analytic evidence likewise indicates that although XAI has expanded across clinical domains, substantial gaps remain in user-centred evaluation, explanation fidelity, methodological transparency, and real-world clinical validation (Abbas *et al.*, 2025).

For a first-aid chatbot, explainability can operate at several levels. First, the system can explain why an accident has been classified in a particular category. Second, it can explain why a particular first-aid action is recommended. Third, it can explain why escalation to professional care is necessary.

For instance, instead of simply stating: "Seek emergency medical attention.", the system could state: "Urgent medical attention is recommended because the reported injury involves loss of consciousness, which may indicate a potentially serious condition."

Such explanations may help users understand the recommendation rather than blindly following an AI-generated instruction.

2.7 Explainability, Trust, and Usability in Healthcare AI

Explainability is not synonymous with simply providing more information. An explanation must be understandable, relevant, sufficiently concise, and related to the decision being communicated. Excessively technical explanations may be inappropriate for ordinary users, particularly during an emergency. A recent systematic review of explanation performance in clinical decision-support systems found that XAI evaluation increasingly considers explanation quality, clinical decision-making, user trust, and usability. The review also noted that patients remain comparatively underrepresented in XAI research and that real-world clinical validation remains limited (Roychowdhury *et al.*, 2025). This observation has direct implications for the proposed study. The chatbot should not be evaluated exclusively by computer scientists or solely through model-performance metrics. Healthcare professionals should assess the clinical correctness of recommendations, while intended users should assess comprehension, usability, trust, and perceived usefulness. A human-centred evaluation is therefore required to determine whether explanations actually improve users' understanding and decision-making.

2.8 Safety and Clinical Reliability of AI Chatbots

Safety represents one of the most important issues in healthcare chatbot research. The ability of large language models to generate highly fluent responses can create an impression of authority even when the underlying information is incorrect.

The emergency-care evaluation by Langdorf *et al.* (2024) provides particularly strong evidence of this problem. Their evaluation showed that several AI chatbot responses contained dangerous or inappropriate information, and that advice regarding when to seek urgent care was frequently incomplete.

This finding suggests that a first-aid chatbot requires **safety-by-design** rather than safety as an afterthought.

A safety-oriented architecture should include:

1. **Evidence-based knowledge retrieval** — recommendations should originate from validated first-aid guidelines.
2. **Rule-based constraints** — the system should prevent prohibited or unsafe recommendations.
3. **Emergency escalation** — specific red-flag symptoms should trigger urgent referral.
4. **Uncertainty detection** — the system should ask additional questions where information is insufficient.
5. **Human oversight** — healthcare professionals should validate the knowledge base and high-risk responses.
6. **Clear limitations** — the chatbot should explicitly communicate that it does not replace professional medical assessment.

The AHA/Red Cross guidelines themselves emphasize the principles of beneficence, non-maleficence, autonomy, and recognition of the first-aid provider's limitations (AHA, 2024). These principles can be translated into software requirements for an AI first-aid system.

2.9 Natural Language Processing and Conversational Interaction

Natural Language Processing is central to the proposed chatbot because users will describe accidents using informal language rather than standardized medical terminology.

An NLP component can perform:

- intent classification;
- named-entity recognition;
- symptom extraction;

- accident-type classification;
- severity-indicator identification;
- question generation; and
- response selection.

Application of transformer-based model may be appropriate because transformer architectures have demonstrated strong performance in natural-language understanding. However, model selection will consider not only predictive performance but also computational requirements, interpretability, latency, and the safety implications of errors.

2.10 Existing Research Gap

Although existing studies have investigated healthcare chatbots, generative AI for emergency-care information, and explainable AI for clinical decision support independently, limited research has integrated these technologies into a safety-constrained, explainable conversational system specifically designed to provide first-aid decision support for domestic accidents. Furthermore, there is limited empirical evidence concerning the development and evaluation of such a system within the Nigerian context, particularly among populations in Ebonyi State. Existing evaluations of AI health chatbots have largely focused on response quality or general healthcare applications, with comparatively less emphasis on combining accident classification, severity assessment, evidence-grounded first-aid recommendations, explanation quality, emergency escalation, and user-centred evaluation in a single system. This study therefore seeks to address this gap by developing and evaluating an explainable AI-based chatbot specifically for domestic accident first-aid assistance.

IV. METHODOLOGY

The proposed first aid chatbot is built using an NLP framework, Rasa and deployed locally using the ngrok that hosts a local web server on its own sub-domain. The chatbot utilizes Telegram, a popular messaging platform, as its chat interface to provide easy access to users with the help of BotFather. The dataset for this chatbot would be generated using web scraping and first-aid documents to structure the data.

Rasa creates the chatbot's model trained on a created dataset using a custom pipeline containing different deep learning-based language models. The bot is first deployed temporarily to a Conversation-Driven Development (CDD) tool, to allow user testing using ngrok, a web tool for creating local servers. Rasa interactive is also used to validate the bot's responses to ensure that the correct actions are executed, and from these changes, the bot learns continuously. For permanent deployment, the chatbot will be deployed on a Digital Ocean remote server and accessed through an API on Telegram. Integrating the CI/CD method into our proposed system allows for the automatic upgrade of the chatbot functionality over time.

3.1 System Architecture

The architecture means working of chatbot starting from user requests to the Bot response (figure 1). The Chatbot background process begins with the user's appeal, for example, "What is PTSD?" to the BOT deployed to the messenger system app like Facebook, Telegram, WhatsApp, Website, Slack, etc. or to the device using speech as input like Google Assistant, Amazon Alexa, Amazon echo dot. After receiving the user's request, the Natural Language Understanding (NLUs) component analyzes it or maps it to the user's intention and, consequently, gathers further related information (intent: "translate," entities: [word: "PTSD"]). Once a CHATBOT reaches the high-level interpretation or confidence score, it must decide how to further proceed and respond accordingly. It can act directly on new information, recall what it has understood, and wait to see what happens next, require more contextual information, or seek clarification (Adamopoulou, Eleni, and Lefteris Moussiades, 2020). For example, "User request to book a Train ticket from Delhi to Mumbai, but to book a ticket other additional information is also required like date of journey, time for the trip. When there is a clear understanding of the request, execution/further action and retrieval of the information occurs. After retrieving the data, BOT intended to perform the requested actions or retrieves the data of interest from its data sources, a BOT Knowledge Base database, or an API call that access external resources (Adamopoulou, Eleni, and Lefteris Moussiades, 2020). The dialogue Management system keeps the information about all the conversations with the users.

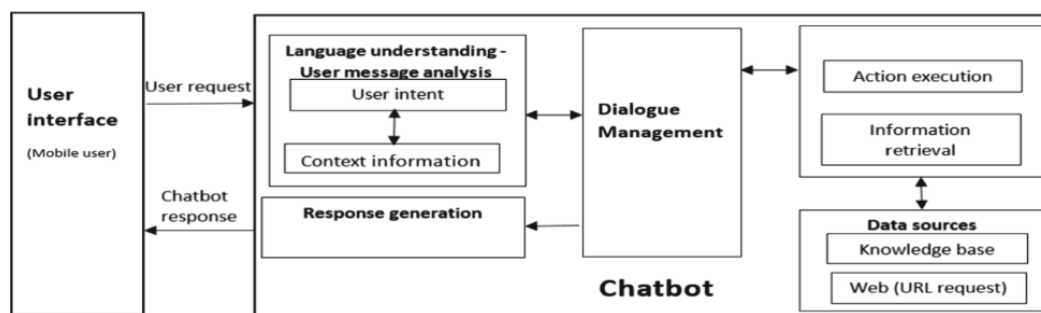


Figure 1: System Architecture

4. Results

4.1 Experimental design

Four classifiers were compared using TF-IDF (Term Frequency-Inverse Document Frequency) word n-gram features: Logistic Regression, Linear Support Vector Machine, Random Forest, and Multinomial Naive Bayes. Four-fold stratified cross-validation was used for the comparative metrics. The SVM was then evaluated on a separate stratified 25% hold-out set for the confusion matrix.

4.2 Model Evaluation Metrics

The models will be evaluated using:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.2 Comparative results

Table 1: Results of the models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Logistic Regression	66.71	61.31	66.07	61.87
SVM	85.96	80.95	86.31	82.10
Random Forest	59.19	52.29	59.52	53.04
Naive Bayes	36.55	28.26	36.31	29.95

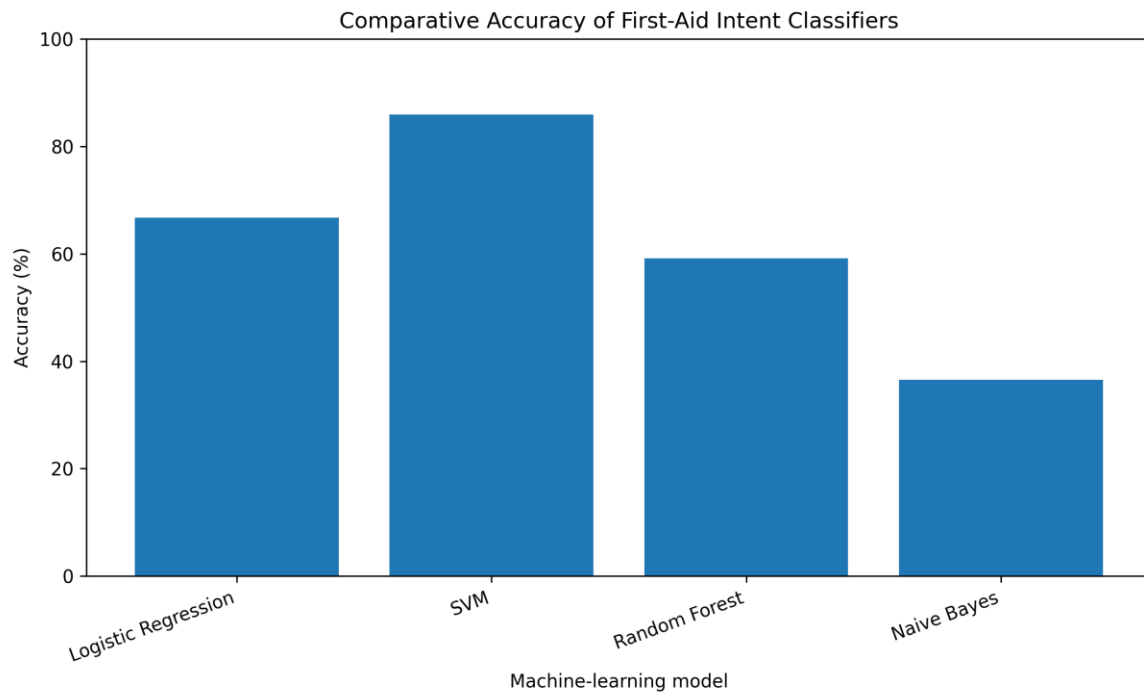


Figure 1: Comparative model accuracy.

The comparative classification accuracy of the four evaluated machine-learning algorithms are shown in the figure 1. SVM achieved the highest mean accuracy of 85.96%, followed by Logistic Regression (66.71%), Random Forest (59.19%), and Multinomial Naive Bayes (36.55%). The superior performance of SVM suggests that it was better able to discriminate among the textual patterns representing the first-aid intent categories in the experimental dataset. This finding is consistent with the suitability of margin-based classifiers for sparse, high-dimensional text representations. Nevertheless, overall accuracy alone is insufficient for evaluating a safety-sensitive first-aid system; consequently, precision, recall, F1-score, and class-level performance were also examined.

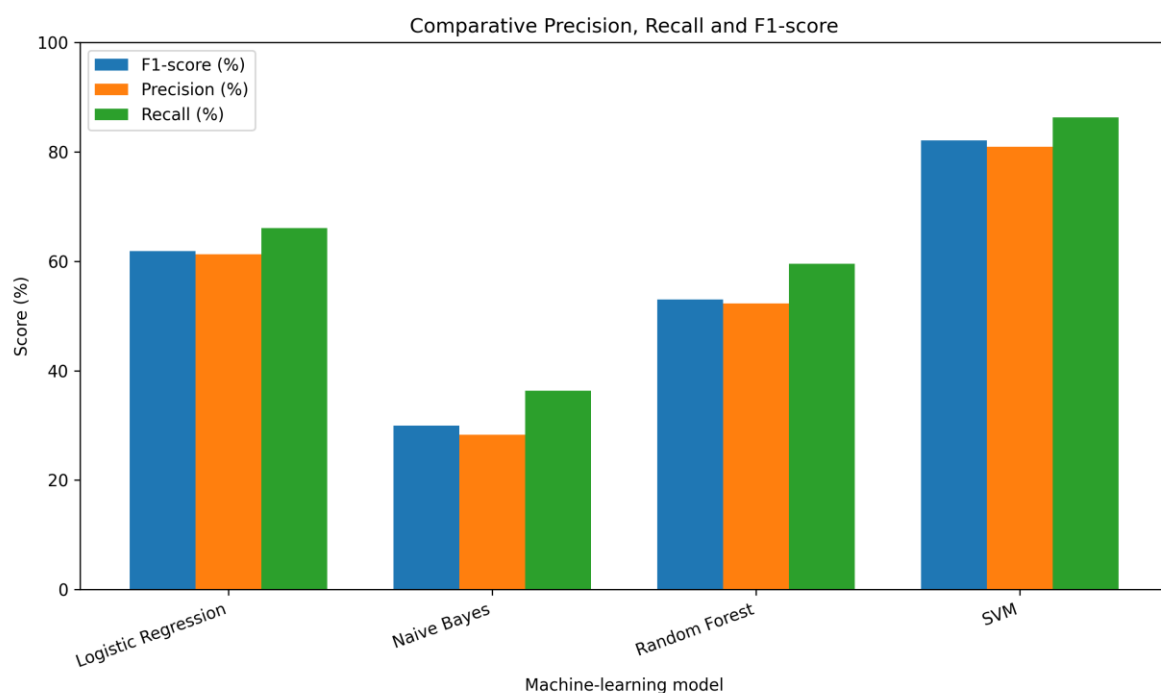


Figure 2: Comparative precision, recall and F1-score.

Figure 2 demonstrates that SVM consistently outperformed the other models across precision, recall, and F1-score. SVM achieved a precision of 80.95%, recall of 86.31%, and F1-score of 82.10%. The relatively high recall is particularly relevant to first-aid decision support because failure to correctly identify an emergency-related intent may result in inappropriate guidance or delayed referral. The results therefore indicate that SVM provides the strongest baseline classification performance among the evaluated models.

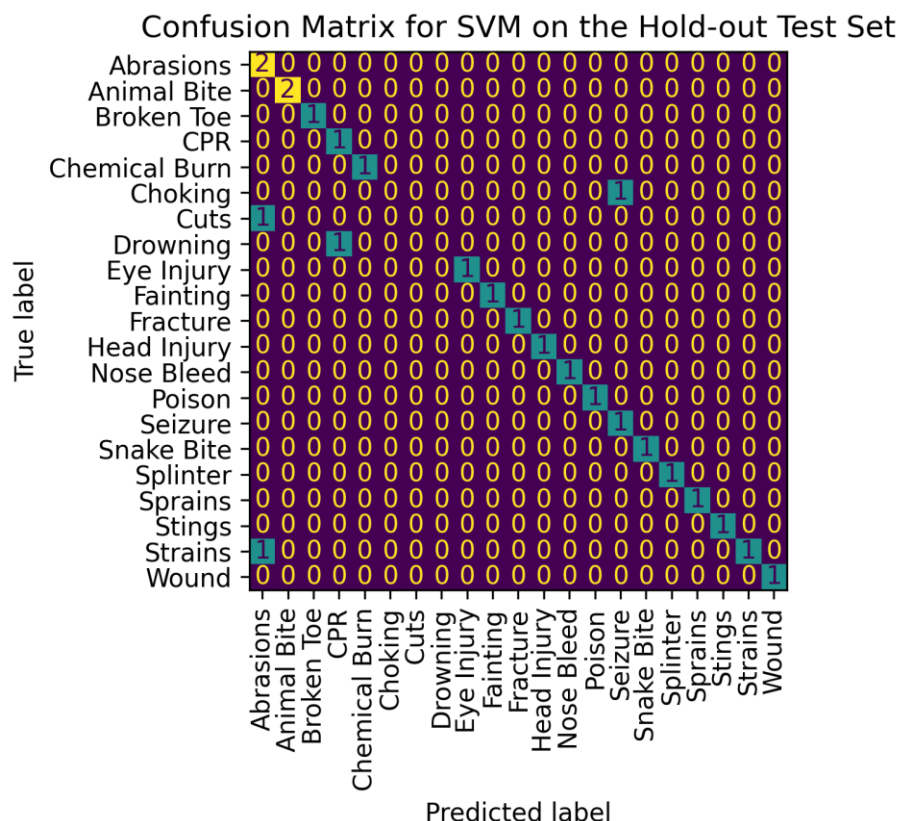


Figure 3. Confusion matrix for the SVM classifier on the hold-out test set.

Figure 3 above presents the confusion matrix of the SVM classifier on the hold-out test set. The diagonal elements represent correctly classified first-aid intents, whereas the off-diagonal elements indicate misclassification between accident categories. The matrix provides a more granular assessment of model performance than overall accuracy because it identifies specific categories in which the classifier encounters difficulty. Particular attention should be given to misclassification involving potentially life-threatening conditions such as choking, poisoning, severe bleeding, drowning, and head injury. Such errors require additional investigation before deployment because incorrect classification may result in inappropriate first-aid recommendations or failure to trigger emergency referral.

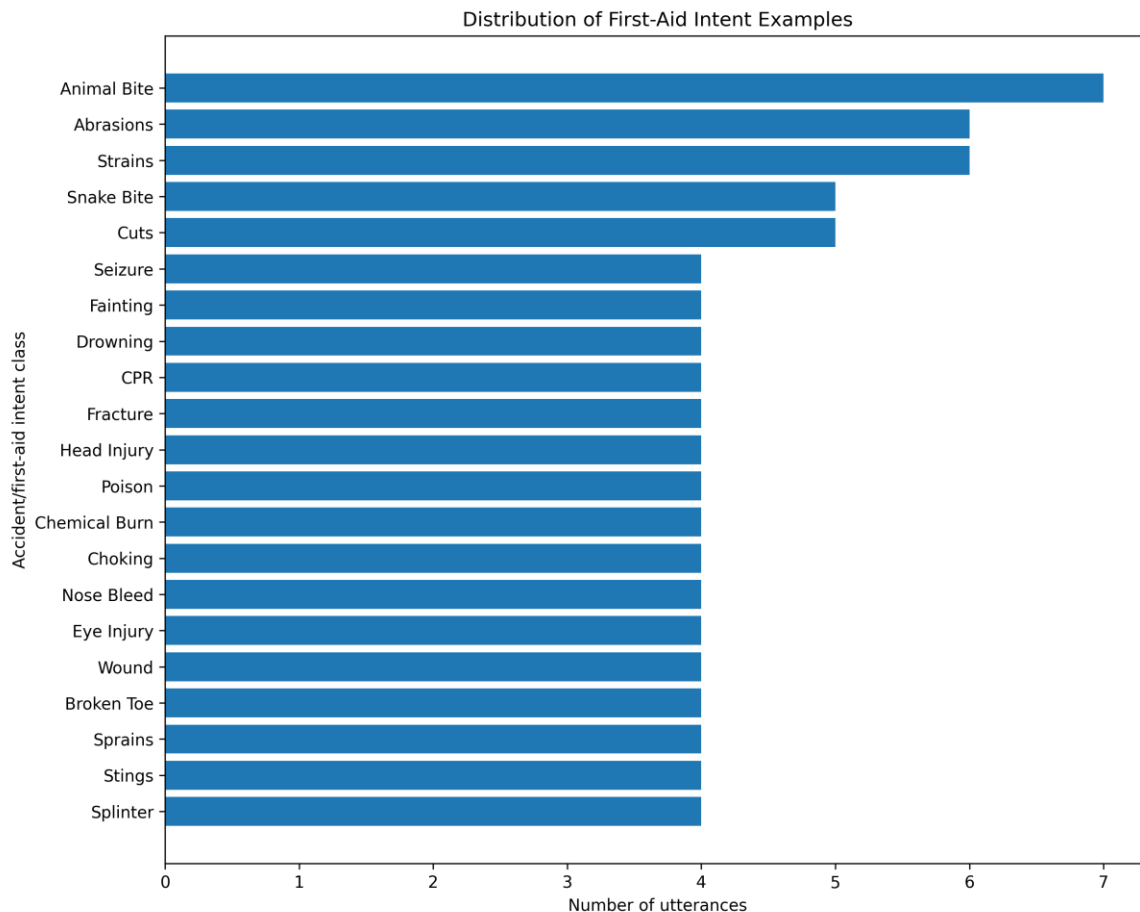


Figure 4. Distribution of intent examples used in the experiment.

The distribution of utterances across the first-aid intent categories used in the experiment as illustrated in figure 4. The distribution demonstrates the relatively small and uneven nature of the experimental dataset. Such class distribution characteristics may influence classifier performance because models generally learn more effectively from classes represented by a larger number of diverse examples. Consequently, the observed performance should be interpreted as preliminary. Future development of the proposed chatbot should incorporate a larger, balanced, and independently validated dataset containing naturally expressed domestic accident scenarios from the target population.

V. Conclusion

This study examined the development of an explainable artificial intelligence (AI)-based chatbot for providing first-aid assistance in domestic accident scenarios among Ebonyi State indigenes. The study was motivated by the need for timely, accessible, and understandable first-aid information, particularly during the critical period between the occurrence of a domestic accident and access to professional healthcare services. The system combines natural language processing, machine learning, evidence-based first-aid knowledge, severity assessment, emergency escalation, and explainable AI to support users in identifying accident scenarios and receiving appropriate initial guidance.

A comparative machine-learning experiment was conducted using a publicly available first-aid intent dataset. Logistic Regression, Random Forest, Support Vector Machine (SVM), and Multinomial Naive Bayes classifiers were evaluated using TF-IDF-based textual representations. The experimental results showed that SVM achieved the best overall performance, recording a mean accuracy of 85.96%, precision of 80.95%, recall of 86.31%, and F1-score of 82.10% under stratified cross-validation. SVM also achieved 83.33% accuracy on the independent hold-out test set. These results indicate that SVM provides a promising baseline for classifying natural-language first-aid queries and identifying the accident intent required by the chatbot's decision-support component.

The incorporation of explainable AI improved the transparency of the chatbot by enabling the system to communicate the basis for its classification and first-aid recommendations. Rather than presenting users with unexplained AI-generated responses, the system can provide understandable reasons for its recommendations and

clearly indicate circumstances requiring immediate medical attention. This approach is particularly important in first-aid applications, where incorrect or misunderstood recommendations may have serious consequences. Overall, the study demonstrates the feasibility of integrating machine learning and explainable AI into a conversational first-aid decision-support system. The approach adopted has the potential to improve access to timely first-aid information and support informed initial responses to domestic accidents. However, the chatbot will serve as a decision support rather than a replacement for medical care.

References

- [1]. American Red Cross guidelines for first aid. *Circulation*, 150(24), e519–e579. <https://doi.org/10.1161/CIR.0000000000001281>
- [2]. Aziz, N. A., Manzoor, A., Qureshi, M. D. M., Qureshi, M. A., & Rashwan, W. (2024). *Explainable AI in healthcare: Systematic review of clinical decision support systems*. *medRxiv*. <https://doi.org/10.1101/2024.08.10.24311735>
- [3]. Birkun, A. A., & Gautam, A. (2024). Large language model-based chatbot as a source of advice on first aid in heart attack. *Current Problems in Cardiology*, 49(1, Pt. A), 102048. <https://doi.org/10.1016/j.cpcardiol.2023.102048>
- [4]. Buowari, D. Y., & Ikpae, B. E. (2022). Domestic accidents among children: A review of cases and prevention strategies. *Nigerian Journal of Medicine*, 63(6), 474–479.
- [5]. Eke, C. S., & Shuib, L. (2025). The role of explainability and transparency in fostering trust in AI healthcare systems: A systematic literature review, open issues and potential solutions. *Neural Computing and Applications*, 37, 1999–2034. <https://doi.org/10.1007/s00521-024-10868-x>
- [6]. Grassini, S., et al. (2025). A systematic review of chatbots in inclusive healthcare: Insights from the last 5 years. *Universal Access in the Information Society*, 24, 195–203. <https://doi.org/10.1007/s10209-024-01118-x>
- [7]. Hewett Brumberg, E. K., Douma, M. J., Alibertis, K., Charlton, N. P., Goldman, M. P., Harper-Kirksey, K., Hawkins, S. C., Hoover, A. V., Kule, A., Leichtle, S., McClure, S. F., Wang, G. S., Whelchel, M., White, L., & Lavonas, E. J. (2024). 2024 American Heart Association and
- [8]. Hua, Y., Xia, W., Bates, D., Hartstein, G. L., Kim, H. T., Li, M., Nelson, B. W., Stromeyer, C. I., King, D., Suh, J., Zhou, L., & Torous, J. (2025). Standardizing and scaffolding health care AI-chatbot evaluation: Systematic review. *JMIR AI*, 4, e69006. <https://doi.org/10.2196/69006>
- [9]. Huo, B., Boyle, A., Marfo, N., McKechnie, D. G. J., Lee, J. S., Mayol, A., Antoniou, S. A., Thirunavukarasu, A. J., & others. (2025). Large language models for chatbot health advice studies: A systematic review. *JAMA Network Open*, 8(2), e2457879. <https://doi.org/10.1001/jamanetworkopen.2024.57879>
- [10]. Kim, J., et al. (2024). XAI-based clinical decision support systems: A systematic review. *Applied Sciences*, 14(15), 6638. <https://doi.org/10.3390/app14156638>
- [11]. Njoku, C. H., & Ekumankama, U. I. (2020). Epidemiology of fractures in a tertiary health institution in South-East Nigeria. *Nigerian Journal of Surgery*. https://doi.org/10.4103/njs.NJS_42_19
- [12]. World Health Organization, & International Committee of the Red Cross. (2018). *Basic emergency care: Approach to the acutely ill and injured*. Geneva, Switzerland: World Health Organization.
- [13]. Yau, J. Y.-S., Saadat, S., Hsu, E., Murphy, L. S.-L., Roh, J. S., Suchard, J., Tapia, A., Wiechmann, W., & Langdorf, M. I. (2024). Accuracy of prospective assessments of 4 large language model chatbot responses to patient questions about emergency care: Experimental comparative study. *Journal of Medical Internet Research*, 26, e60291. <https://doi.org/10.2196/60291>