Simulation of a hybrid medical decision support system for interpreting medical tests

Kouamé Appoh[1], Yapi Debarros Emmannuella[2], Kra Lagasane Ouattara[2*], Brou Konan Marcellin[1*]
[1]Institut National Polytechnique Félix Houpouet-Boigny, Yamoussoukro, Côte d'Ivoire
[2]Department of Mathematical Computer Science, Alassane Ouattara University, Bouake, Côte d'Ivoire

Abstract
This article presents the design, simulation, and evaluation of a hybrid medical decision support system for interpreting biological tests, particularly in the diagnosis of ischemic stroke. The model combines a predictive machine learning module (Random Forest) and a deductive expert module based on validated medical rules, linked by a weighting parameter α that modulates their respective influence. Experiments conducted on a dataset of 5,110 observations from the Kaggle platform show that optimal performance is achieved for $0.6 \leq \alpha \leq 0.8$, corresponding to an AUC of 0.820. This range reflects a balance between generalization capacity and clinical interpretability, validating the value of hybridization between symbolic reasoning and statistical learning. The results demonstrate the relevance of this approach for reducing diagnostic errors, ensuring the traceability of medical decisions, and adapting to resource-limited contexts in Africa. The proposed model provides a basis for the development of intelligent decision support systems that are explainable and integrable into clinical practice.
Keywords : expert system, hybrid system, ischemic stroke, machine learning, mathematical model, medical decision, simulation.

1. Introduction

Medical decision-making is a complex process requiring a combination of expert clinical judgment and objective analysis of biological data. In many developing countries, particularly in Africa, the lack of specialists, the multiplicity of chronic diseases, and the growing volume of medical data complicate the reliability of diagnoses. The integration of machine learning into medical practice has transformed diagnostic and predictive approaches [9], providing tools capable of analyzing vast data sets and generating personalized recommendations. However, as Luo et al. (2023) point out, clinical confidence in artificial intelligence systems depends heavily on their explainability and transparency of reasoning [7]. In this context, medical decision support systems (MDSS) offer an innovative solution to support practitioners by combining symbolic reasoning (expert systems) and machine learning approaches. However, each of these paradigms has its limitations: purely expert models lack adaptability, while predictive models are often opaque and difficult to explain. Hybrid models therefore appear to be a promising way of reconciling algorithmic performance and clinical interpretability. This work follows on from A.C. Aka et al. (2025), who proposed a theoretical framework for coupling symbolic reasoning and machine learning [5]. The objective of this study is to experimentally validate this model through the simulation of a hybrid medical decision support system applied to the assessment of ischemic stroke risk. More specifically, the aim is to implement a coupled deductive-predictive system based on a set of medical rules and a machine learning model, to analyze the influence of the coupling parameter α on overall performance, and to demonstrate the clinical and methodological relevance of such a system in an African context. The rest of the thesis is structured as follows: the second chapter presents an analysis of existing work, the third chapter describes the simulation methodology, the fourth chapter presents the results and discussion, and finally, the fifth chapter offers a summary conclusion of the work accomplished.

In the field of medical decision support systems, hybrid approaches reconcile algorithmic performance and clinical interpretability [4], offering a promising solution for effectively integrating machine learning with expert reasoning. While exploring the scientific literature, we identified and catalogued the following systems:

The authors of [3] proposed one of the first hybrid systems applied to breast cancer diagnosis. Their model integrates data mining and a fuzzy neural network optimized by a genetic algorithm, demonstrating synergy between adaptive learning and symbolic reasoning. However, the complexity of the architecture and the high computational cost limited its clinical applicability.

The author of [2] laid the theoretical foundations for clinical decision support systems, emphasizing the integration of medical knowledge and dynamic learning. His work paved the way for robust models but remains theoretical without experimental validation.

The authors of [1] developed a decision support system based on case-based reasoning (CBR) and a multi-criteria approach. Their MCDS model enabled learning from real cases while considering several decision criteria. However, the lack of an automatic adaptation mechanism made it difficult to update the system in response to new clinical cases.

The authors of [10] designed a system based on a hybrid data mining approach combining Naive Bayes and J48. The combination improved the reliability of diagnosis based on symptoms. However, the data sample was limited and clinical tests were absent, reducing the external validity of the model.

The authors of [11] proposed a hybrid clinical system for managing mineral disorders related to chronic kidney disease. Although the model showed 78.27% agreement with clinical practice, generalization remained limited by sample size and dependence on specific cases.

The authors of [5] introduced a mathematical model coupling symbolic reasoning and machine learning. This work laid the foundations for a rigorous analytical framework, but its scope remained theoretical without full experimental implementation.

2. Materials and Methods

Following on from the work of [5], which introduced a mathematical model for coupling symbolic reasoning and machine learning, this study proposes an experimental implementation aimed at empirically validating the relevance of the initial theoretical framework. The methodology is based on the characterization of the dataset, the description of the development tools, and the presentation of the hybrid system simulation algorithm. This logic of coupling symbolic rules and machine learning is consistent with the work of [8].

a) Characteristics of the dataset

The deductive module is based on 24 medical rules developed from personal research and validated by a physician to ensure their relevance and clinical legitimacy. These rules take into account the major risk factors for ischemic stroke: age, hypertension, heart disease, glucose levels, body mass index (BMI), and smoking.

Examples of medical rules:

- Purpose: To provide a reliable and interpretable deductive framework, where each decision can be traced and justified.
- Characteristics: Deterministic and explainable, each patient is assigned a probability and confidence level based on the rules that are activated.

The predictive module is based on a structured dataset from Kaggle, used to train a Random Forest model for predicting stroke risk. This choice is justified by its robustness in the face of noisy data, its ability to model complex interactions between variables, and its balance between performance and readability. It provides a relevant methodological starting point before exploring more advanced algorithms.

Dataset characteristics:

- Size: 5,110 records, ensuring a sufficient statistical basis for learning.
- Main variables:
  - Demographic: age
  - Clinical: hypertension, heart disease, blood sugar, BMI
  - Lifestyle: smoking
- Target variable: stroke (1 = patient who has had a stroke; 0 = no stroke)

Objective : To provide the predictive module with representative data enabling it to learn the complex relationships between clinical variables and stroke risk.

b) Tools and development environment

Several tools were used to simulate the hybrid model in order to ensure the flexibility, performance, and reproducibility of the analyses:

- Python: Main language of the project, chosen for its versatility and rich scientific libraries.
- Scikit-learn: Implementation of the Random Forest model and calculation of performance indicators (Accuracy, F1-score, AUC, Brier score).
- Pandas / NumPy: Data manipulation, cleaning, and structuring.
- Matplotlib / Seaborn: Visualization of the performance and distribution of the α parameter.
- Pathlib: File path management and automation of result saving.

All of these tools have made it possible to create a reproducible and robust simulation environment, conducive to experimentation with the hybrid system and in-depth analysis of the results.

c) Presentation of the simulation algorithm

The hybrid system simulation algorithm consists of several successive steps :

- Data preprocessing:

- Cleaning (filling in missing values, particularly BMI, using the median).
- Encoding categorical variables using the one-hot encoding method.
- Separation into training (75%) and test (25%) sets, stratified according to the target variable (stroke).

- Implementation of the machine learning module:

Based on the dataset, the Random Forest model produces a stroke risk probability (proba_ML) and a confidence measure (C_ML) calculated according to the distance to 0.5.

- Implementation of the expert module:

The expert system's inference engine (forward chaining) generates a deductive probability (proba_EX) and a confidence level (C_EX) based on logical medical rules using the same clinical variables.

- Module coupling:

The two outputs are combined according to the weighting parameter α: $Y\_final = \alpha \times \hat{Y}\_ML + (1 - \alpha) \times \hat{Y}\_EX$.
Different strategies for calculating α were tested:
- Fixed α (constant value, e.g., 0.5).
- Expert enriched by ML (low α, except in cases of expert uncertainty).
- Expert-guided ML (strong α, except in cases of rule violations).
- Bidirectional coupling (α varies according to the confidence levels of the two modules).

• Performance analysis:

As there is no single metric for evaluating a medical model, several complementary indicators were used:
- Accuracy: proportion of correct predictions, overall performance measure.
- F1-score: harmonic mean between precision and recall, essential for rare diseases.
- AUC-ROC: ability of the model to distinguish positive from negative cases across all possible thresholds.
- Brier score: measure of the calibration of predicted probabilities, indicator of clinical reliability.

The combined use of these measures ensures a balanced assessment of the model's discriminatory ability, its clinical sensitivity, and the consistency of its probabilistic predictions. Thus, after describing the data, tools, and algorithm, the following section presents the performance evaluation of the hybrid system and an analysis of its relevance in a medical context.

4. Results and Discussion

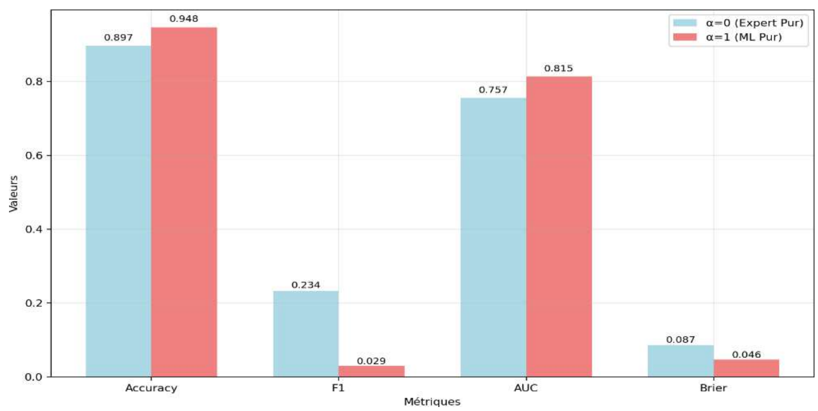

Figure 1 : Sensitivity analysis of the performance of the hybrid model to the ML weighting (α)

Performance metrics highlight the system's non-linear behavior in response to variations in the α parameter. Accuracy and AUC reach their maximum levels when α is in the range [0.6; 0.8], indicating that a compromise where machine learning slightly dominates human expertise optimizes the model's discrimination capacity and overall performance. This optimal zone highlights that machine learning, when guided by expert knowledge, can effectively exploit interactions between variables while benefiting from the decision-making stability provided by clinical rules.

However, analysis of other indicators reveals a structural tension between clinical and statistical objectives. The F1 score peaks when α = 0, confirming that the expert system performs better in detecting positive cases, which is essential in a medical context

where diagnostic sensitivity may be a priority. Conversely, the Brier Score reaches its minimum for α = 1, showing that the pure machine learning model offers better probability calibration.

This confirms that in the absence of explicit expertise, statistical learning optimizes probabilistic consistency but may lose accuracy in critical or rare cases. Thus, these results illustrate a compromise inherent in decision support systems. medical: reconciling the discriminating capacity of ML, the clinical sensitivity of the expert system, and the reliability of probabilistic calibration. The optimum observed around α ≈ 0.7 validates the idea that synergy between these approaches is preferable to the isolated use of any one of them.

In this context, the choice of the α parameter is not solely a matter of numerical optimization, but also a methodological decision focused on clinical requirements.

An α parameter too close to 1 would essentially favor statistical performance at the expense of clinical caution, while an α that is too low would risk limiting the model's ability to adapt its predictions to complex or atypical cases. It therefore seems necessary to adopt an intermediate value that guarantees accuracy, generalization, and clinical reliability, while maintaining interpretability, which is an essential criterion in medical environments.

Looking ahead, the adjustment of α constitutes a first approach to hybridization. Beyond this linear weighting, different coupling strategies can be explored:

• sequential integration (the expert validates or completes the ML decision),

• probabilistic fusion models,

• dynamic rules depending on the type of clinical case,
• or even more advanced neuro-symbolic architectures that explicitly incorporate medical knowledge into neural networks.

These avenues open up the possibility of designing hybrid systems.

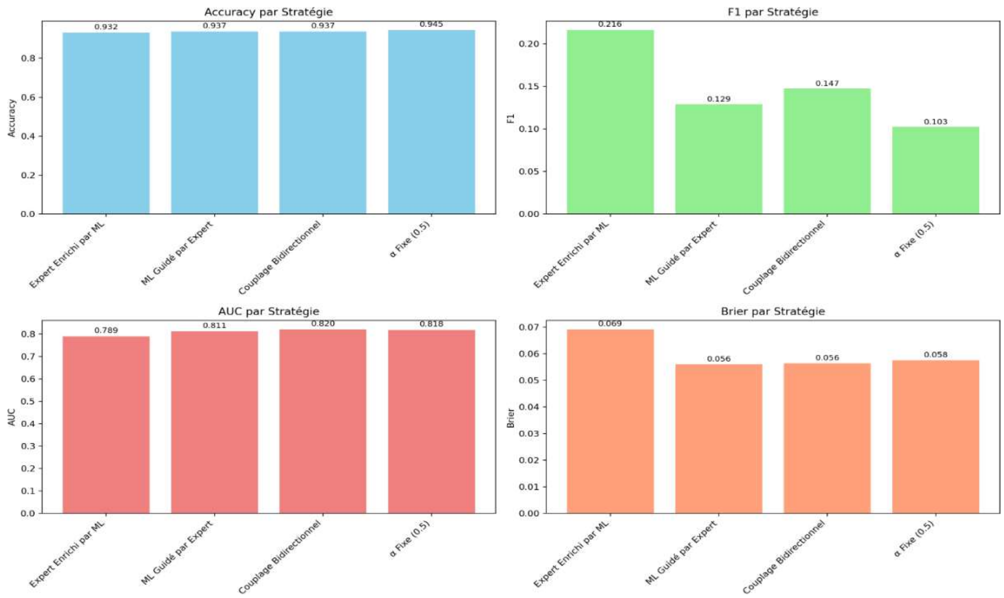

Figure 2 : Comparison of clinical decision support strategies

All of the strategies tested showed high overall accuracy, with accuracy consistently above 0.93. This level of performance demonstrates the overall robustness of the approaches studied.

However, it is important to note that accuracy, although informative, is not sufficient on its own to judge the actual performance of a diagnostic system, particularly in a medical context where classes may be unbalanced or positive cases rare. From this perspective, the “ML-Enriched Expert” strategy stands out clearly: it simultaneously achieves the best F1 and AUC scores, demonstrating an enhanced ability to correctly identify positive cases and effectively discriminate between sick and healthy patients.

These results confirm that combining explicit expert reasoning with insights from machine learning not only improves the model's discriminatory power, but also optimizes its ability to capture subtle clinical patterns that expert rules alone might overlook.

However, this strategy has a higher Brier score, reflecting slightly less accurate probabilistic calibration. In other words, although it excels at detecting positive cases, the probabilities it assigns may sometimes be less well aligned with the actual probabilities of disease. This observation illustrates a classic challenge in medical decision-making systems: striking a balance between discrimination, clinical sensitivity, and reliable probability calibration. While accuracy remains useful for assessing overall performance, it can mask critical errors in identifying positive cases. In the context of diagnosis, particularly for rare diseases, metrics such as AUC and F1-Score appear more appropriate. The F1-Score reflects the model's ability to recognize patients who are actually affected, while the AUC measures the ability to separate classes in various decision threshold contexts.

Thus, the superiority of the “ML-Enriched Expert” strategy on these two indicators attests to its relevance as a hybrid approach. It combines the structural robustness and clinical consistency of expert reasoning with the ability of machine learning to generalize and process complex data. The slight deterioration in Brier should be interpreted not as a major weakness but as a natural trade-off between calibration and detection capability, which is often tolerable in early screening where sensitivity is prioritized.

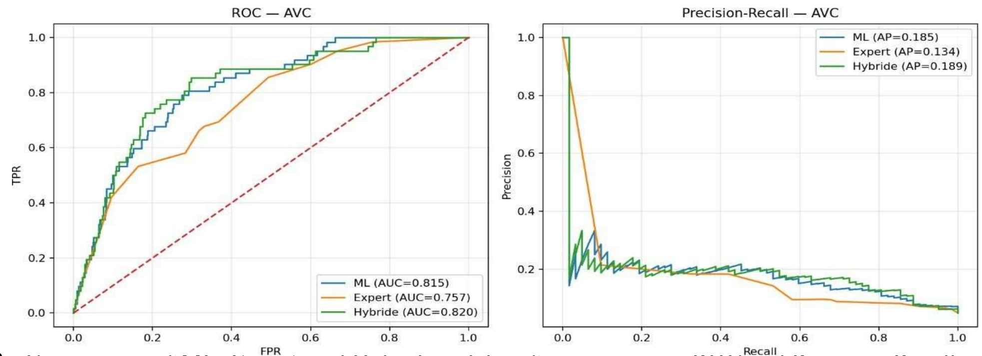

Figure 3 : Comparison of ML, Expert, and Hybrid model performance using ROC and Precision-Recall curves

Examination of the ROC curve shows that the hybrid model performs best overall with an AUC of 0.820, slightly ahead of the pure Machine Learning model (AUC = 0.815) and significantly outperforming the expert system (AUC = 0.757). This superiority reflects the hybrid model's increased ability to differentiate positive cases from negative cases across all possible decision thresholds.

This trend is corroborated by the Precision-Recall curve, which is particularly appropriate for evaluating methods in the context of imbalanced classes or rare diseases. Indeed, the Average Precision (AP) is 0.189 for the hybrid model, compared to 0.185 for the Machine Learning model and 0.134 for the expert model, proving that the mixed approach is the most effective for identifying patients who are actually ill while maintaining a high level of accuracy.

In the medical diagnostics sector, it is essential to simultaneously reduce false negatives so as not to overlook cases of illness, and false positives in order to avoid unnecessary treatment.

The AUC, by evaluating overall discrimination ability, provides a comprehensive assessment of model performance. However, the Precision-Recall curve is particularly crucial in contexts where disease prevalence is low, as it evaluates the model's ability to maintain adequate performance when detecting rare cases. The conclusions drawn indicate that the hybrid model effectively combines the strengths of explicit clinical reasoning and machine learning, thus providing more accurate identification of positive cases, even when they are underrepresented. In addition, the α parameter appears to play a crucial role in optimizing the system. By adjusting the respective contributions of the expert module and the ML module, it allows for the determination of an optimal range, between 0.6 and 0.8, where performance is maximized.

Furthermore, the importance of the α parameter is crucial in optimizing the system. By adjusting the respective contributions of the expert module and the ML module, it is possible to determine an optimal range, between 0.6 and 0.8, where performance is maximized. This zone illustrates a delicate balance where the rigor and clinical consistency derived from expert rules combine with the ML's ability to learn from huge volumes of data and identify complex patterns. Hybridization is therefore not only an arrangement, but also a first-rate strategy capable of surpassing the approaches examined individually.

In the medical field, this structure is a crucial advantage. It maintains the traceability, interpretability, and security of expert systems while taking advantage of predictive power where clinical expertise alone may not suffice. This model serves as a robust decision-making tool for healthcare professionals: it supports medical expertise without supplanting it, while offering additional assurance in difficult diagnostic cases. From a scientific perspective, these results validate the relevance of hybrid systems in medicine by proving that the combination of structured medical knowledge and machine learning algorithms leads to more stable and comprehensive performance. This evidence-based demonstration paves the way for future advances in both the improvement of hybrid architectures and their implementation in other clinical fields. It contributes to the advent of AI-enhanced medicine that is more accurate, safer, and more in tune with contemporary requirements, where human intelligence and computational capabilities complement each other for the benefit of the patient in order to meet current standards.

The results obtained confirm the robustness and consistency of the hybrid model in predicting stroke risk.

The analysis reveals an optimal performance range for α values between 0.6 and 0.8, where the system achieves a balance between statistical discrimination and clinical relevance. This nonlinear behavior is due to the fact that the model's accuracy is not linear with respect to the number of features used. The analysis reveals an optimal performance zone for α values between 0.6 and 0.8, where the system achieves a balance between statistical discrimination and clinical relevance. This non-linear behavior illustrates that machine learning, when guided and supervised by an expert knowledge base, effectively exploits the complex interactions between variables while maintaining the stability and interpretability necessary for medical practice.

However, the differences observed between the metrics show that there is still tension between statistical accuracy and clinical sensitivity: the expert module remains superior for detecting positive cases, while the pure machine learning model stands out for its better probabilistic calibration. The hybrid system successfully reconciles these two dimensions, offering an optimal compromise where the complementary nature of the approaches improves the

overall quality of the diagnosis. The “ML-enriched expert” strategy thus appears to be the most balanced, combining the rigor of symbolic reasoning with the adaptive capacity of machine learning, while maintaining a high level of overall performance (AUC = 0.820).
These results constitute an experimental validation of the mathematical model proposed by A.C. Aka et al. (2025) and pave the way for the exploration of more dynamic forms of hybridization, such as probabilistic fusion or neuro-symbolic architectures, which are likely to further increase the reliability and personalization of medical diagnosis. They are also in line with the observations of Krause et al. (2020), according to which predictive performance must always be accompanied by reliable clinical interpretation.

6. Conclusion

The results obtained confirm the robustness and consistency of the hybrid model in predicting stroke risk. The analysis highlights an optimal performance zone for α values between 0.6 and 0.8, where the system achieves a balance between statistical discrimination and clinical relevance. This nonlinear behavior demonstrates that machine learning, when guided and supervised by an expert knowledge base, effectively exploits complex interactions between variables while maintaining the stability and interpretability necessary for medical practice. However, the differences observed between the metrics show that there remains a structural tension between statistical accuracy and clinical sensitivity: the expert module remains superior for detecting positive cases, while the pure machine learning model stands out for its better probabilistic calibration. The hybrid system thus manages to reconcile these two dimensions, offering an optimal compromise where the complementarity of the approaches improves the overall quality of the diagnosis. The “ML-enriched expert” strategy appears to be the most balanced, combining the rigor of symbolic reasoning with the adaptive capacity of machine learning, while maintaining a high level of overall performance (AUC = 0.820). These results constitute experimental validation of the mathematical model proposed by A.C. Aka et al. (2025) and pave the way for exploring more dynamic forms of hybridization, such as probabilistic fusion or neuro-symbolic architectures, in order to further increase the reliability and personalization of medical diagnosis.

Conflicts of Interest

The authors declare no conflict of interest with respect to the publication of this article.

Références

1.Abdelhak M., Baghdad A. et al. (2013). A Hybrid Decision Support System: Application on Healthcare.
2.Berner E. (2009). Clinical Decision Support Systems: Theory and Practice. Springer.
3. Da Silva I. G. L., Amorim B. P. et al. (2002). Integration of Data Mining and Hybrid Expert System.
4.Holzinger, A. et al. (2021). Toward Explainable AI in Healthcare: The Hybrid Approach. Journal of Biomedical Informatics, 118, 103792.
5. A.C. Aka et al. (2025). Un système hybride d’aide à la décision pour la médecine moderne : modèle mathématique et validation théorique.
6.Krause, J. et al. (2020). Interpretable Machine Learning Models for Predictive Healthcare. Nature Machine Intelligence, 2(6), 338–347.
7.Luo, Y. et al. (2023). Explainable AI for Healthcare: Hybrid Neuro-Symbolic Systems and Trustworthy Decision-Making. Artificial Intelligence in Medicine, 142, 102626.
8.Nguyen, P., et al. (2022). Combining Rule-Based Reasoning and Deep Learning for Clinical Decision Support Systems. Frontiers in Artificial Intelligence, 5, 839614.
9.Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. New England Journal of Medicine, 380(14), 1347–1358.
10.Suhas A. B., Sumukha J. S. et al. (2019). Medical Expert System using Data Mining and Machine Learning.
11.Syed I. A., Su W. J. et al. (2021). Clinical Decision Support System Based on Hybrid Knowledge Modeling.