Towards a Hybrid AI Solution for Emergency Department Patient Triage

Georgia Sowerby[1], Ol'Tunde Ashaolu[2], Radu Calinescu[1] and Stephen Connor[3]

[1]Department of Computer Science, University of York, York, United Kingdom
[2]York and Scarborough Teaching Hospitals NHS Foundation Trust, Scarborough, United Kingdom
[3]Department of Mathematics, University of York, York, United Kingdom

Abstract

Hospital Emergency Departments (EDs) worldwide are overwhelmed. The number of people attending EDs increases yearly; however, there is a shortage of doctors and many unfilled vacancies. Unfortunately, this means waiting times can be very long for patients, and doctors are overworked. This paper summarises a PhD project to develop and integrate two key components of an end-to-end hybrid Artificial Intelligence (AI) solution for ED patient triage, building on an existing system DAISY. The first component involves using formally verified machine learning (ML) models to classify patient characteristics that play a major role in ED triage but are difficult to determine by directly asking patients. The second component is a Bayesian network that takes the predictions of these verified ML models and additional data obtained from the patient (e.g., medical history, current symptoms, and vital signs) and generates an advisory patient triage report supporting the decision-making of ED medical teams.

Keywords

Clinical decision support system, Bayesian network, Rule-based expert system, Machine learning

1. Introduction

Emergency departments worldwide face growing challenges due to increasing demand and limited resources. In the UK, the National Health Service (NHS) Constitution handbook pledges a maximum four-hour ED waiting time, with an operational standard that 95% of patients are admitted, transferred, or discharged within this timeframe [1]. Yet in 2022-23, NHS reports show that out of 25 million ED patients, 30% waited longer than four hours to receive care [2]. This highlights strain on the system and raises concerns about timely and effective care. Furthermore, British Medical Association data reveal a significant shortage of doctors in England, and many vacant positions [3].

These challenges are not unique to the UK. The World Health Organisation has reported that many European countries are facing "substantial shortages and gaps" [4]. The culmination of these pressures results in a challenging and high-stress working environment for ED staff, many of whom work long hours and report increasing levels of job dissatisfaction [5].

A key component in managing ED patient flow is *triage*, the *"clinical process to prioritise patients, completed before a full assessment to support effective management of demand and flow, identifying time critical requirements for patients"* [6]. This process typically comprises five stages [7]:

1. **Reception:** Administrative staff gather preliminary information by observing the patient and listening to their concerns. Patients who need immediate care can be escalated at this stage; otherwise, they progress to the next.
2. **History and Symptoms:** A triage clinician collects detailed patient information, asking questions about their medical history and current symptoms.

RuleML+RR'25: Companion Proceedings of the 9th International Joint Conference on Rules and Reasoning, September 22–24, 2025, Istanbul, Turkiye

georgia.sowerby@york.ac.uk (G. Sowerby); tunde.ashaolu@nhs.net (O. Ashaolu); radu.calinescu@york.ac.uk (R. Calinescu); stephen.connor@york.ac.uk (S. Connor)
0000-0002-6145-8336 (G. Sowerby); 0000-0002-7829-4884 (O. Ashaolu); 0000-0002-2678-9260 (R. Calinescu); 0000-0002-9785-2159 (S. Connor)

3. **Vital Sign Measurement:** Often occurring in tandem with the previous stage, the clinician measures and records the patient's vital parameters (e.g., temperature, heart rate, respiratory rate, blood pressure, oxygen saturation).
4. **Initial Assessment:** The clinician analyses the collected data to determine the patient's triage score and suggest potential assessments. This analysis can lead to actions such as escalating the case, returning the patient to the waiting room, or transferring them to a different department.
5. **Senior Clinician Review:** A senior clinician reviews the collected patient information and any potential assessments suggested by the clinician from Stage 4, then conducts a physical examination. They then decide on a treatment plan, which might include treating and discharging, referring for further investigations, admitting, or transferring to another facility.

This process is prone to inconsistency, especially during busy periods, and automating it is challenging as automated triage systems often rely solely on structured data (i.e., medical history, patient-reported symptoms, and vital signs). In contrast, doctors also use physical examinations and clinical intuition to assess clinically relevant patient characteristics (i.e., observable signs that indicate a patient's condition). Automated systems currently lack sensory input and therefore miss this important information. However, cameras and microphones can capture visual and auditory cues that may offer additional insight into a patient's condition. Incorporating this data into triage systems could improve diagnostic accuracy. ML models are well suited to learning such patterns, and formally verifying these models can help improve confidence in their predictions, which is essential in clinical environments.

Clinical decision support systems (CDSSs) offer one possible solution to improve triage consistency and support overloaded clinicians by helping doctors make fair, evidence-based decisions regarding patient care. They are classified into knowledge-based or non-knowledge-based [8]. Knowledge-based systems use explicit rules, usually formulated as *if-then* statements, to represent medical knowledge. In contrast, non-knowledge based systems use ML techniques to identify patterns from large datasets. Although increasingly common in research, the real-world use of non-knowledge based CDSSs is limited due to concerns including explainability and limited access to high-quality data [8]. Knowledge-based systems are often preferred in clinical environments because their recommendations can be traced to clear, interpretable rules. However, these systems are deterministic and do not model uncertainty, which can limit their flexibility when information is incomplete or ambiguous.

One prototypical example of such systems is DAISY (Diagnostic AI System for Robot-Assisted A&E Triage) [7, 9, 10, 11], a knowledge-based CDSS developed to support ED triage by automating Stages 2 through 4 of the process outlined earlier. It uses a rule-based architecture that prioritises transparency and explainability, but it does not model uncertainty or provide probability-ranked outputs.

To address DAISY's limitations and to lay the foundation for its next generation, this research proposes a hybrid CDSS that combines machine learning, probabilistic reasoning, and rule-based logic. The overarching aim of this research is to employ AI technologies to develop a CDSS that automates Stages 2 – 4 (and possibly part of Stage 5) of ED triage with accurate results, reducing ED patient waiting times and ED medical staff workload. To that end, the following research questions guide this work:

- RQ1 How can a Bayesian network component be used alongside or instead of the current DAISY rule-based expert system to improve the accuracy of patient triage and assessment?
- RQ2 Which patient characteristics relevant to ED triage can be identified with acceptable accuracy using verified ML?
- RQ3 How can the ML classifiers from RQ2 be integrated with the Bayesian network from RQ1 to further improve triage and assessment in DAISY?

2. Related Work

Many clinical decision support systems have been proposed over the years. One of the earliest and most influential was INTERNIST-I, developed in the 1970s to address the growing complexity of internal medicine [12]. At its peak, the INTERNIST-I knowledge base included 572 diagnoses and over 4,000 patient findings, with more than 4,000 rules [13]. The rules were written by medical experts using textbooks and their own experience. However, the system had several limitations: it treated users as passive, required highly specific terminology for input, and generated consultations that lasted up to 75 minutes, which made it impractical for fast-paced clinical settings such as the ED.

To address these issues, Quick Medical Reference (QMR) was developed in the early 1980s. Like INTERNIST-I, QMR used the same knowledge base but acted more as an interactive information tool. It allowed users to input findings and receive feedback, supported by a completer feature to improve usability. Around the same time, ILIAD was introduced. While it began as a deterministic system, it later adopted a Bayesian network formalism. The resulting model included 11,406 nodes, with some structures extending to 36 levels and common findings shared by up to 62 parent nodes [14].

The transition to Bayesian networks in ILIAD reflected their advantages in handling uncertainty, a key challenge in medical reasoning [15, 16]. Bayesian networks also provide a clear graphical structure and can model causal relationships, making them well-suited to safety-critical applications like ED triage. In a comparative study based in the ED, ILIAD outperformed QMR, providing correct diagnoses in 72% of cases versus QMR's 52%, although both systems generated long differential lists [17].

ML has been used in emergency medicine applications such as predicting hospital admission [18, 19], workflow optimisation [20, 21], critical care [22, 23], and specific conditions such as sepsis or stroke [24, 25]. However, the lack of explainability in many of these models raises concerns [26, 8]. To address these concerns, recent research has focused on translating rule-based expert systems into Bayesian networks (e.g., [27, 28]).

ML has been used to detect clinically relevant patient characteristics from visual and auditory data. For example, there are several approaches to cough detection and classification using audio recordings, with many models able to distinguish between wet and dry coughs, as well as coughs associated with specific conditions such as Covid-19 or pertussis [29, 30, 31]. Similarly, jaundice has been detected from images of the sclera and skin using models trained to either predict levels of bilirubin, an orange-yellow pigment formed by the breakdown of red blood cells, or to classify images as jaundiced or non-jaundiced [32, 33, 34]. Much of the existing work in this area focusses on neonatal jaundice, with limited application to adult patients. In most cases, a colour reference square is included in the images to support colour normalisation during preprocessing. Facial droop, such as droop on one side of the face, has also been detected using ML models trained on facial images [35, 36, 37]. Common approaches typically include face detection, facial landmark extraction, and the measurement of distances between key points to identify asymmetry.

While ML classifiers have been developed for a range of individual clinically relevant patient characteristics, we are not aware of any existing work that combines them within a single expert system to support clinical decision-making. To address this gap, we build upon DAISY, which was selected as the foundation for this research because it is actively maintained and specifically tailored to emergency triage. As members of the DAISY team, we have direct access to the codebase and to the clinical collaborators involved in its development. This makes it a more suitable platform than larger but less accessible systems such as INTERNIST-I.

3. Approach

3.1. Background

DAISY [7, 9, 10], the CDSS on which our research is focused, gathers four categories of medically relevant information for an ED patient being triaged:

- **Demography:** The patient's medical history
- **Anatomical:** Specific parts of the body affected.
- **Subjective:** Symptoms reported directly by the patient.
- **Objective:** The patient's vital signs.

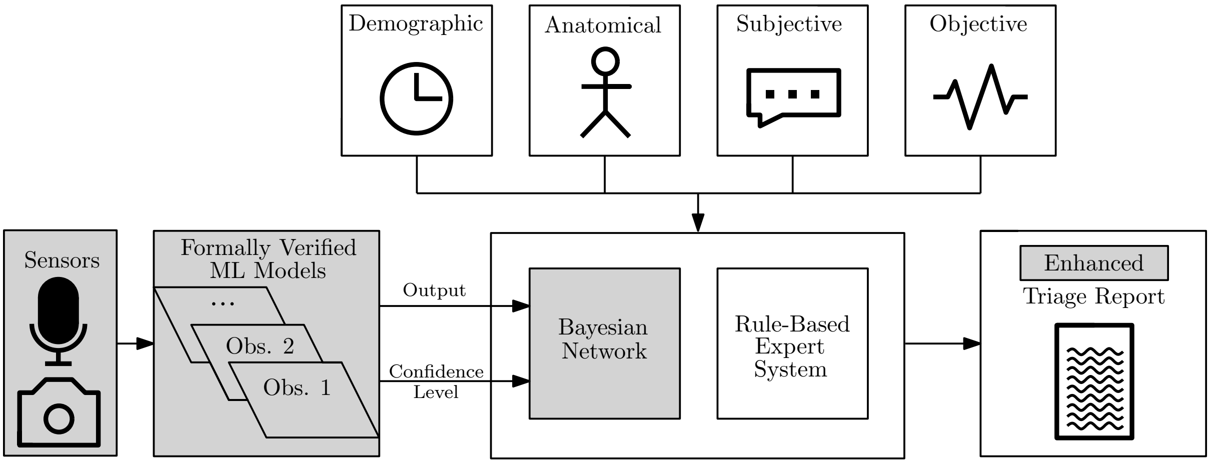

Figure 1: Preliminary architecture of a hybrid AI solution for ED triage and assessment, with new contributions shaded.

Each category includes a set of variables that collectively describe the patient's clinical presentation in a structured manner. For example, the Demography category includes variables such as age, sex, and recent trauma. The Anatomical category includes variables that identify affected body parts such as head and chest. The Subjective category includes variables representing self-reported symptoms including pain and nausea. The Objective category includes variables for vital signs such as pulse rate and temperature.

The values of the variables within the Demography, Anatomical, and Subjective categories are obtained by asking the patient questions, which they answer via a touch-screen interface. The Objective category is obtained by instructing the patient to use medical devices in the room to measure their vital signs.

This information is processed by dAvInci, the rule-based expert system at the core of the DAISY project. dAvInci currently uses over 200 doctor-specified rules to output the patient's triage score and a set of potential assessments, suggested investigations, treatments and referrals. A triage report, which contains the information gathered by DAISY alongside the output from dAvInci, is made available to the patient's doctor.

DAISY is not merely conceptual. It has been implemented in a clinical setting and is currently undergoing a feasibility study at Scarborough Hospital as part of a registered clinical trial [11]. The study will involve 100 patients and will evaluate DAISY's acceptability, consultation duration, patient engagement, and clinical concordance compared to standard triage. These measures will provide insight into how well DAISY fits into real-world practice and how its outputs align with clinician assessments. The study will also generate real-world triage data to support the evaluation of the Bayesian network approach developed in this research.

3.2. Overview

Figure 1 presents the preliminary architecture of the proposed hybrid AI solution for ED triage and assessment. Non-shaded components represent existing DAISY functionality, while shaded components represent the new contributions developed in this work. These contributions are organised into two main components:

1. **Component 1 (Verified ML models).** Addressing RQ2, this component enables DAISY to capture clinically-relevant patient characteristics, allowing future variants of the system to mimic aspects of clinical intuition.
2. **Component 2 (Bayesian network).** Addressing RQ1, this component augments DAISY with probabilistic reasoning, allowing the system to represent uncertainty and generate outputs such

as probability-ranked assessments, with a view to better support clinical decision-making.

The integration of these two components (RQ3) extends the existing DAISY framework by combining probabilistic reasoning with clinically relevant patient characteristics.

3.3. Component 1: Verified ML for Patient Characteristics

Doctors routinely use all their senses when assessing patients, drawing on years of experience to make quick, intuitive judgements. This 'clinical intuition' often involves the subconscious recognition of subtle but clinically significant characteristics and plays an important role in patient assessment.

The enhanced DAISY aims to mimic this capability by using formally verified ML models to classify patient characteristics captured by external sensors, such as cameras and microphones. Each classifier outputs both a prediction and an associated confidence score, which can then be used to augment the structured data captured by DAISY, leading to improved decision-making.

3.4. Component 2: Bayesian Network for Probabilistic Reasoning

To supplement DAISY's deterministic reasoning, we introduce a Bayesian network. Bayesian networks are particularly well-suited to healthcare applications, as they capture both causal relationships between symptoms and diseases and the uncertainty inherent in medical reasoning. This is important because, while patterns do exist, no two patients present in exactly the same way.

An additional strength of Bayesian networks is their graphical structure, which makes the reasoning process more interpretable. In a safety-critical domain such as emergency medicine, explainability is essential because clinicians must be able to understand and trust the system's outputs.

By incorporating and quantifying uncertainty in this way, potential assessments can be ordered according to their probability, thereby significantly enhancing the triage reports currently produced by DAISY.

3.4.1. Single Malady Example

To illustrate this approach, we use 'SIRS; Meningitis' as a running example.[1] It is one of over 200 maladies in DAISY's knowledge base.

Example: SIRS; Meningitis

Consider a rule for identifying possible SIRS; Meningitis. The rule is triggered if the following conditions are met:

- **Demography** The patient has no history of recent physical trauma
- **Anatomy and Subjective** The patient reports a problem with their head and is bothered by bright lights (photophobia)
- **Objective** At least two of the following are true: abnormal temperature (e.g., low or high), elevated respiratory rate, elevated pulse rate

This rule can be formalised as:

$$
\begin{aligned}
&\texttt{Demography\_RecentTrauma} = \texttt{no} \\
&\wedge\ \texttt{Head\_BotheredByBrightLights} = \texttt{yes} \\
&\wedge\ \Big((\texttt{Objective\_Temperature} \neq \texttt{normal} \wedge \texttt{Objective\_RespiratoryRate} = \texttt{high}) \\
&\qquad \vee\ (\texttt{Objective\_Temperature} \neq \texttt{normal} \wedge \texttt{Objective\_PulseRate} = \texttt{high}) \\
&\qquad \vee\ (\texttt{Objective\_RespiratoryRate} = \texttt{high} \wedge \texttt{Objective\_PulseRate} = \texttt{high})\Big) \\
&\implies \texttt{Malady\_SIRSMeningitis} = \texttt{yes}
\end{aligned}
$$

[1]We use the label 'SIRS; Meningitis' to refer to cases where systemic inflammatory response syndrome (SIRS) is present in conjunction with or as a result of meningitis.

Our aim is to encode this logic within a Bayesian network, using probabilities to relax the original strict binary rules. Rather than requiring at least two abnormal vital signs to consider SIRS; Meningitis as a potential assessment, the probability should increase with each abnormal vital sign. This provides a more flexible and realistic way to support clinical decision-making under uncertainty.

3.5. Integration

The integration of verified ML classifiers with the Bayesian network enables the enhanced DAISY system to incorporate clinically relevant patient characteristics into its probabilistic reasoning. Each classifier detects a specific characteristic using sensor data and outputs both a prediction and confidence score.

For each characteristic, two nodes are added to the Bayesian network: one representing the classifier's prediction, and the other its degree of confidence in that output. These nodes are connected directly to the relevant malady nodes, with the direction of the link going from the malady to the classifier nodes.

The classifier outputs and their confidences are then used as evidence during inference, allowing them to influence the probabilities of associated maladies. For example, if a jaundice classifier returns a positive result with high confidence, the probability of related conditions such as liver disease or pancreatitis increases. If the confidence is low, the effect on the probability is lesser.

As part of the integration process, synthetic data must be generated to train the updated Bayesian network and evaluate its performance. This is non-trivial and will require close collaboration with medical experts to ensure the data is realistic.

4. Methodology

The methodology for the delivery of the approach summarised in the previous section is presented below on a per component basis.

4.1. Verified ML for Patient Characteristics

A review of relevant ML and clinical research literature was conducted, and an interview with an emergency medicine consultant helped identify a set of clinically relevant characteristics. As detailed in Section 5.1, one characteristic was selected from this set of shortlisted characteristics based on its clinical usefulness, detection feasibility, and suitability for data collection.

For the characteristic used in this study, no suitable open-access dataset was available, so a clinical study was initiated to collect the required data. If additional characteristics were to be incorporated, appropriate datasets would need to be identified or new data collected as necessary. Data augmentation techniques were applied to increase dataset size where appropriate.

A suitable formally verified ML model will be created and validated for the chosen characteristic.

4.2. Bayesian Network for Probabilistic Reasoning

The structure of the Bayesian network is derived from DAISY's existing expert rule base. Each variable appearing in one of the DAISY categories Demography, Objective and Malady is mapped to a corresponding node in the Bayesian network. Symptoms are always recorded with a corresponding anatomical variable in DAISY, and these are combined into a single node type Anatomy_Subjective in the Bayesian network. DAISY also includes a special Anatomy variable 'General', to indicate that a symptom is not localised to a specific anatomical location (e.g. nausea).

Directed edges are added to the network based on predefined causal assumptions representing clinical knowledge:

- **Demography → Malady** A patient's demographic and medical history can influence the probability of developing certain conditions.

- **Malady → Anatomy_Subjective, Objective** Once present, a malady is expected to cause symptoms and physiological changes.

Nodes from Demography, Anatomy_Subjective, and Objective may be connected to multiple Malady nodes, reflecting the fact that a single variable (e.g. temperature) can be associated with several possible conditions. The Bayesian network also supports directed links between Malady nodes themselves, capturing relationships where one condition may influence the likelihood or severity of another. This represents a key extension beyond the current DAISY system, which does not account for any co-morbidities. We are currently exploring options for which co-morbidities to include in the Bayesian network.

Continuous variables such as temperature and respiratory rate are discretised into clinically meaningful intervals, based on thresholds defined in DAISY's existing rule base. Although not reported here, we are also experimenting with representing these variables as continuous nodes within the Bayesian network to allow for more flexible modelling.

The network construction was implemented using the Bayes Server [38] Java API within R. The structure of the DAISY rule base permitted the majority of the network topology to be generated programmatically in a systematic and efficient manner.

4.2.1. Single Malady Example

For our running example, SIRS; Meningitis, we constructed the structure of the Bayesian network from the corresponding rules in DAISY's knowledge base. As no suitable real-world dataset was available, we generated synthetic data for 1000 patients using an R script. The distributions were designed to reflect realistic ED conditions and to conform to DAISY's original rule constraints. The dataset was split into 700 records for training and 300 for testing, and training was performed using Bayes Server's Relevance Tree algorithm, chosen for its efficiency and compatibility with discrete and discretised variables. This single-malady example is intended to demonstrate the feasibility of network construction and parameter learning, while the full Bayesian network will ultimately be trained on real-world ED triage data, such as that currently being collected as part of a clinical trial for DAISY.

Descriptions of the variables used in this example are provided in Table 1. Continuous variables were discretised into clinically meaningful intervals. For example, temperature was divided into five states (e.g., Very Low [0, 35], Low [35.1, 36], etc.), respiratory rate into five states, and pulse rate into six states.

5. Preliminary Results

5.1. Verified ML for Patient Characteristics

The review of the clinical research literature (e.g., [39, 40, 41, 42, 43]) led to ten broad categories of clinically-relevant patient characteristics being identified:

- Skin (condition, hydration)
- Gait and Movement
- Respiration (rate, depth, coughing, choking, wheezing)
- Behaviour (emotional state, communication)
- Level of Consciousness
- Facial Features and Expressions
- **Eyes** (**sclera colour**, symmetry, blinking, pupil size, reaction)
- Personal Hygiene, Grooming and Dress (e.g., cleanliness)
- Position, Posture, Height and Build
- Visible Issues (bleeding, bruising, physical abnormalities)

Table 1
Variables used in the SIRS; Meningitis example

Variable	Type	Description	Distribution
Malady_SIRSMeningitis	Binary	Indicates whether the individual has SIRS; meningitis	Probability of ”Yes” = 5%
Demography_RecentTrauma	Binary	Indicates whether the individual experienced recent physical trauma	When Malady_SIRSMeningitis = ”Yes”, probability of ”Yes” = 15% When Malady_SIRSMeningitis = ”No”, probability of ”Yes” = 40%
Head_BotheredByBrightLights	Binary	Indicates whether the individual reported sensitivity to bright lights	When Malady_SIRSMeningitis = ”Yes”, probability of ”Yes” = 88% When Malady_SIRSMeningitis = ”No”, probability of ”Yes” = 15%
Objective_Temperature	Continuous	Body temperature (°C)	When Malady_SIRSMeningitis = ”Yes”, Objective_Temperature is drawn randomly from either $N(35.5, 0.25)$ or $N(38.5, 0.35)$ with equal probability, bounded to [34, 41]. When Malady_SIRSMeningitis = ”No”, Objective_Temperature is drawn from $N(36.8, 0.7)$, bounded to [34, 41].
Objective_RespiratoryRate	Continuous	Respiratory rate (breaths per minute)	When Malady_SIRSMeningitis = ”Yes”, Objective_RespiratoryRate is drawn from $N(26, 7)$, bounded to [6, 35]. When Malady_SIRSMeningitis = ”No”, Objective_RespiratoryRate is drawn from $N(16, 7)$, bounded to [6, 35].
Objective_PulseRate	Continuous	Pulse rate (beats per minute)	When Malady_SIRSMeningitis = ”Yes”, Objective_PulseRate is drawn from $N(120, 35)$, bounded to [30, 180]. When Malady_SIRSMeningitis = ”No”, Objective_PulseRate is drawn from $N(75, 25)$, bounded to [30, 180].

From this list, a yellow tint to the sclera (the whites of the eyes) was selected as the first characteristic for development of an ML-detection component. This decision was informed by both the clinical literature review and discussions with medical experts, with the intention to expand to additional characteristics in future work.

Jaundice, often characterised by yellowing of the skin and sclera, occurs due to elevated levels of bilirubin in the blood. It can be a symptom of various underlying conditions, including liver disease, bile duct obstruction, blood disorders, neonatal jaundice, and pancreatitis.

The aim is to develop a verified ML classifier that categorises images of the sclera into one of four classes: Yellow, Borderline, Not Yellow, Inconclusive.

As no suitable open-source datasets exist, a clinical study called *Measuring Icterus from Sclera Tinge Yellowness (MiSTY)* is being conducted to collect eye images and corresponding blood bilirubin levels. This dataset will support the development and validation of a verified ML classifier for integration into the hybrid DAISY system.

5.2. Bayesian Network for Probabilistic Reasoning

A full Bayesian network structure for the DAISY knowledge base has been created, consisting of 304 nodes and over 1,500 links. This network has not yet been trained.

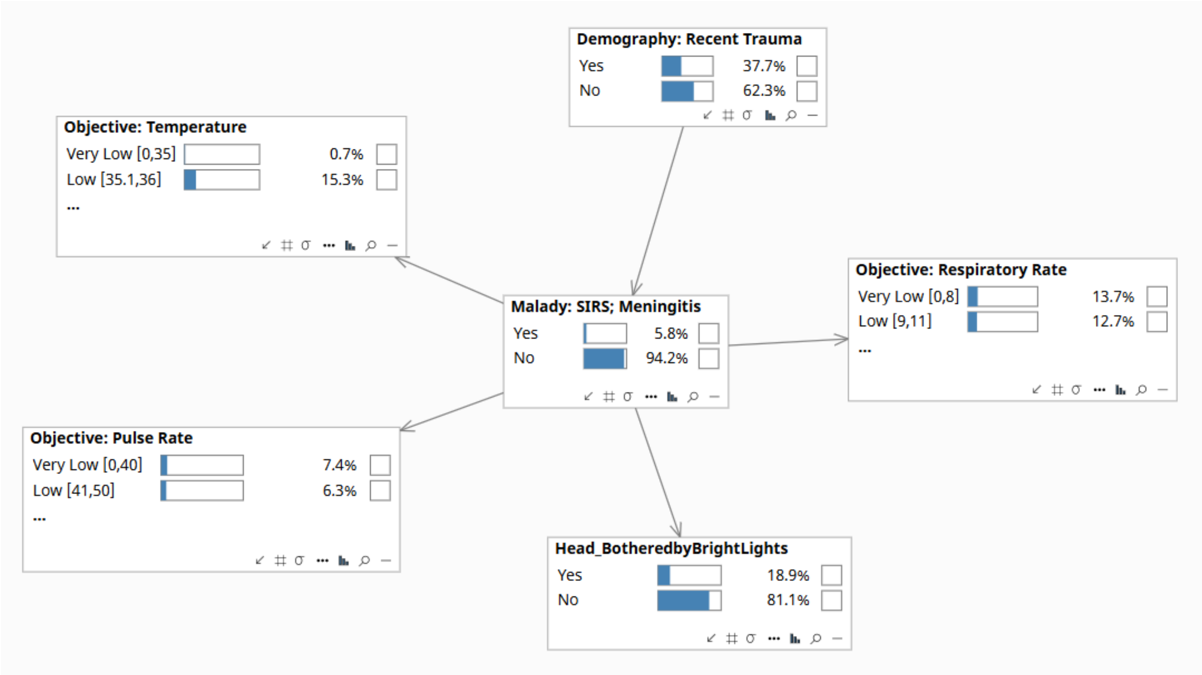

Figure 2: Pictorial representation of the trained Bayesian network for SIRS; Meningitis, as displayed in Bayes Server.

5.2.1. Single Malady Example

The trained Bayesian network for the SIRS; Meningitis example was evaluated on the synthetic dataset. The structure of the network was derived from DAISY's rules, while the conditional probability tables were learned from data.

Figure 2 illustrates the resulting Bayesian network. As expected, the model assigns higher probabilities of SIRS; Meningitis when more supporting evidence is present. For example, if a patient has no recent trauma, is bothered by bright lights, and has a high temperature and high pulse rate (but a normal respiratory rate), the model estimates a 76.9% chance that the patient has SIRS; Meningitis. If the respiratory rate is also high, this probability increases to 93.5%.

This example highlights the model's ability to handle variation in clinical presentation, recognising that while specific conditions are typically associated with certain patterns, patients with the same diagnosis may differ in their medical history, affected anatomy, reported symptoms, and vital signs.

Figure 3 presents box plots of the predicted probabilities for SIRS; Meningitis, grouped by the actual class label ("Yes" or "No"). The top plot shows results on the training set, and the bottom plot shows the test set. In both cases, the predicted probabilities are clearly separated between positive and negative cases. Patients in the "Yes" group consistently receive higher predicted probabilities than those in the "No" group, suggesting that the model has learned to distinguish between the two classes effectively.

The similarity between training and test results suggests that the model generalises well without signs of overfitting.

6. Conclusion

This paper described a PhD project aiming to develop a hybrid AI solution for ED triage. The proposed approach extends the DAISY system by introducing formally verified ML models to recognise clinically relevant patient characteristics and a Bayesian network to support probabilistic reasoning.

So far, the project has defined a preliminary system architecture, identified potential patient characteristics to be modelled using ML, selected one characteristic, submitted a clinical study for ethical review

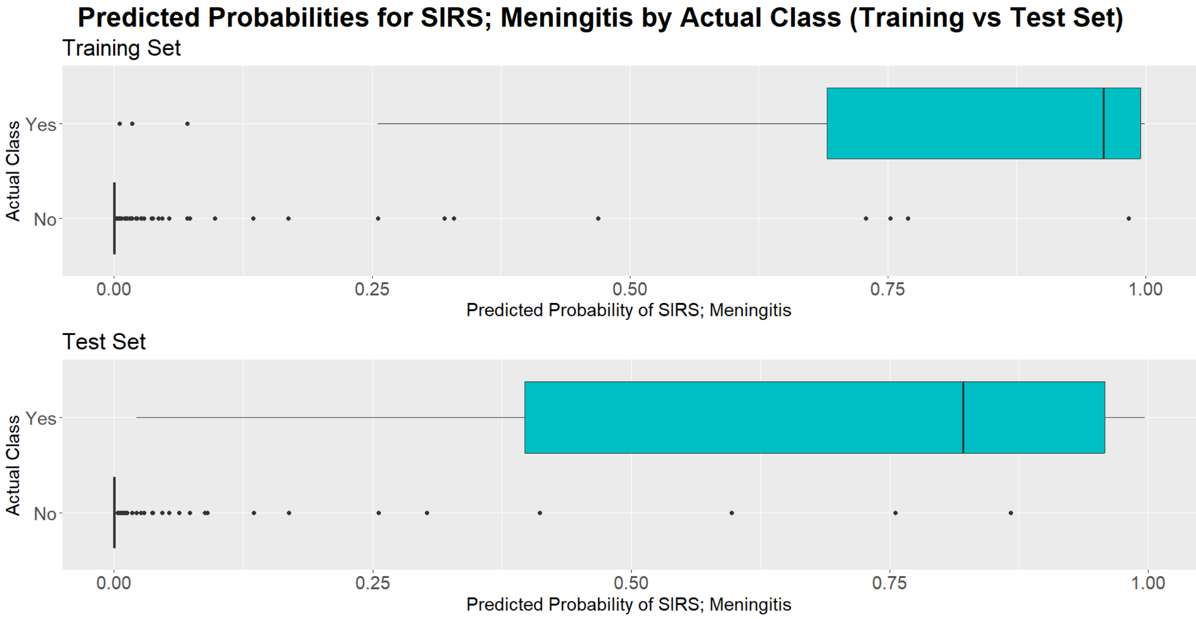

Figure 3: Predicted probabilities for SIRS; Meningitis by actual class. Top: Training Set; Bottom: Test Set

to provide training data, and created the structure of the Bayesian network. A single malady example has been used to demonstrate the feasibility of applying Bayesian networks within this approach.

Next steps include developing and validating the verified ML model for the selected characteristic using data from the clinical study, training the full Bayesian network, and integrating these components. The performance of the enhanced DAISY system will then be evaluated. Future work will also consider advancing probabilistic reasoning under uncertainty by incorporating continuous vital signs and modelling co-morbidities.

Acknowledgments

This research is funded by a UKRI Trustworthy Autonomous Node in Resilience Doctoral Training Programme studentship.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Nuffield Trust, A&E waiting times, 2024. URL: https://www.nuffieldtrust.org.uk/resource/a-e-waiting-times.

[2] NHS Digital, Hospital Accident & Emergency Activity, 2022-23, 2023. URL: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-accident--emergency-activity/2022-23.

[3] British Medical Association, NHS medical staffing data analysis, 2024. URL: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/workforce/nhs-medical-staffing-data-analysis.

[4] World Health Organization, Ticking timebomb: Without immediate action, health and care workforce gaps in the European Region could spell disaster, 2022. URL: https://www.who.int/europe/news/item/14-09-2022-ticking-timebomb--without-immediate-action--health-and-care-workforce-gaps-in-the-european-region-could-spell-disaster.

[5] Sartini, Marina and Carbone, Alessio and others, Overcrowding in emergency department: causes, consequences, and solutions — a narrative review, Healthcare 10 (2022) 1625.

[6] NHS England, Guidance for emergency departments: initial assessment, n.d. URL: https://www.england.nhs.uk/guidance-for-emergency-departments-initial-assessment/.

[7] O. Ashaolu, W. Lyons, et al., Autonomous Emergency Triage Support System, in: CSCI'23, 2023, pp. 1332–1337. doi:10.1109/CSCI62032.2023.00220.

[8] R. T. Sutton, D. Pincock, et al., An overview of clinical decision support systems: benefits, risks, and strategies for success, NPJ Digital Medicine 3 (2020) 17.

[9] University of York, The DAISY project: Diagnostic AI System for Robot-Assisted A&E Triage, 2023. URL: https://www.cs.york.ac.uk/research/projects/daisy-project/.

[10] B. A. Townsend, K. L. Plant, et al., Medical practitioner perspectives on AI in emergency triage, Frontiers in Digital Health 5 (2023) 1297073.

[11] York Teaching Hospitals NHS Foundation Trust, Daisy - diagnostic ai system for robotic and automated triage and assessment (daisy), ClinicalTrials.gov ID NCT06571838, 2025. URL: https://clinicaltrials.gov/study/NCT06571838.

[12] J. D. Myers, The Background of INTERNIST I and QMR, in: ACM Conference on History of Medical Informatics, 1987, pp. 195–197. doi:10.1145/41526.41543.

[13] R. A. Miller, M. A. McNeil, et al., The INTERNIST-1/QUICK MEDICAL REFERENCE Project—Status Report, Western Journal of Medicine 145 (1986) 816.

[14] Y.-C. Li, et al., Automated transformation of probabilistic knowledge for a medical diagnostic system, Proc Annu Symp Comput Appl Med Care (1994) 765–769.

[15] E. Kyrimi, K. Dube, et al., Bayesian networks in healthcare: What is preventing their adoption?, Artificial Intelligence in Medicine 116 (2021) 102079.

[16] L. Uusitalo, Advantages and challenges of Bayesian networks in environmental modelling, Ecological Modelling 203 (2007) 312–318.

[17] M. A. Graber, D. VanScoy, How well does decision support software perform in the emergency department?, Emergency Medical Journal (2003).

[18] G. Feretzakis, G. Karlis, et al., Using machine learning techniques to predict hospital admission at the emergency department, The Journal of Critical Care Medicine 8 (2022) 107–116.

[19] W. S. Hong, A. D. Haimovich, et al., Predicting hospital admission at emergency department triage using machine learning, PLOS ONE 13 (2018) e0201016.

[20] M. Moustapha Mbaye, F. MH Abu Salem, et al., A new machine learning workflow to create an optimal waiting list in hospitals, in: ICMHI 2023, 2023, pp. 159–163.

[21] J. W. Joseph, E. L. Leventhal, et al., Machine learning methods for predicting patient-level emergency department workload, The Journal of Emergency Medicine 64 (2023) 83–92.

[22] Y. Liu, J. Gao, et al., Development and validation of a practical machine-learning triage algorithm for the detection of patients in need of critical care in the emergency department, Scientific Reports 11 (2021) 24044.

[23] Y. Raita, T. Goto, et al., Emergency department triage prediction of clinical outcomes using machine learning models, Critical Care 23 (2019) 1–13.

[24] T. M. Sullivan, Z. P. Milestone, et al., Development and validation of a Bayesian belief network predicting the probability of blood transfusion after pediatric injury, Journal of Trauma and Acute Care Surgery 94 (2023) 304–311.

[25] S.-F. Sung, L.-C. Hung, et al., Developing a stroke alert trigger for clinical decision support at emergency triage using machine learning, International Journal of Medical Informatics 152 (2021) 104505.

[26] H. Chang, W. C. Cha, Artificial intelligence decision points in an emergency department, Clinical and Experimental Emergency Medicine 9 (2022) 165.

[27] M. Korver, P. J. Lucas, et al., Converting a rule-based expert system into a belief network, Medical Informatics 18 (1993) 219–241.

[28] S. Thirumuruganathan, M. Huber, et al., Building Bayesian Network based expert systems from rules, in: IEEE SMC 2011, 2011, pp. 3002–3008.

[29] L. Kvapilova, et al., Continuous Sound Collection Using Smartphones and Machine Learning to Measure Cough, Digital Biomarkers 3 (2020) 166–175.

[30] M. Pahar, et al., COVID-19 Cough Classification Using Machine Learning and Global Smartphone Recordings, Computers in Biology and Medicine 135 (2021) 104572.

[31] A. Imran, et al., AI4COVID-19: AI Enabled Preliminary Diagnosis for COVID-19 from Cough Samples via an App, Informatics in Medicine Unlocked 20 (2020) 100378.

[32] J. H. Park, et al., Deep-Learning-Based Smartphone Application for Self-Diagnosis of Scleral Jaundice in Patients with Hepatobiliary and Pancreatic Diseases, Journal of Personalized Medicine 11 (2021) 928.

[33] A. Althnian, et al., Neonatal Jaundice Diagnosis Using a Smartphone Camera Based on Eye, Skin, and Fused Features with Transfer Learning, Sensors 21 (2021) 7038.

[34] T. S. Leung, et al., Screening Neonatal Jaundice Based on the Sclera Color of the Eye Using Digital Photography, Biomedical Optics Express 6 (2015) 4529–4538.

[35] C.-Y. Chang, et al., Application of Machine Learning for Facial Stroke Detection, in: 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP), IEEE, 2018, pp. 1–5.

[36] G. S. Parra-Dominguez, et al., Facial Paralysis Detection on Images Using Key Point Analysis, Applied Sciences 11 (2021) 2435.

[37] Lee, Aaron and Che, Daniel and Qi, Kevin and Patton, Evan and Zeng, Tingying, Mobile device application of deep learning for rapid onset stroke detection, AIP Conference Proceedings 2562 (2023) 020006. doi:10.1063/5.0138862.

[38] Bayes Server Ltd., Bayes Server, 2025. URL: https://www.bayesserver.com/.

[39] L. S. Bickley, Bates' Guide to Physical Examination and History Taking, 13th ed., international ed. ed., Lippincott Williams & Wilkins, Philadelphia, 2020.

[40] O. Epstein, et al., Clinical Examination, 4th ed. ed., Mosby, Edinburgh ; New York, 2008.

[41] F. Faustinella, The Power of Observation in Clinical Medicine, International Journal of Medical Education 11 (2020) 250.

[42] I. Adler, The Medical Gap: Intuition in Medicine, Medicine, Health Care and Philosophy 25 (2022) 361–369.

[43] F. W. Gardiner, The Art of Self-Knowledge and Deduction in Clinical Practice, Annals of Medicine and Surgery 10 (2016) 19–21.