International Journal on Computational Modelling Applications
Vol. 01, Iss. 01, S. No. 001, pp. 1-19, July 2024
ISSN (E): 3048-8516

A Study on Anaphylactic-shock Forecasting System with Knowledge-Based Intelligent Architecture

Meghna Sarkar, Siddhartha Chatterjee, Sudipta Hazra, Ritwika Ghosh, Soumi Chakraborty

Department of Information Technology, Kalyani Government Engineering College, Kalyani - 741235, Nadia, West Bengal, India, meghnasarkar00@gmail.com
Department of Computer Science and Engineering, College of Engineering and Management Kolaghat, KTPP Township, Purba Medinipur - 721171, West Bengal, India, siddhartha.chatterjee31@gmail.com
Department of Computer Science and Engineering, Asansol Engineering College, Kanyapur, Asansol - 713305, West Bengal, sudiptahazra.nitdgp@gmail.com
Department of Computer Science and Engineering, Institute of Science and Technology, Paschim Medinipur - 721201, West Bengal, India, ritwikaghosh86@gmail.com
Department of Optometry & Vision Sciences, NSHM Knowledge Campus, Durgapur - 713212, West Bengal, India, soumichakraborty58@gmail.com

How to cite this paper: Meghna Sarkar, Siddhartha Chatterjee, Sudipta Hazra, Ritwika Ghosh, Soumi Chakraborty, "A Study on Anaphylactic-shock Forecasting System with Knowledge-Based Intelligent Architecture" *International Journal on Computational Modelling Applications* , Vol. no. 01, Iss. No 01, S No. 001, pp. 1–19, July 2024.

Received: 02/06/2024
Revised: 30/06/2024
Accepted: 15/07/2024
Published: 31/07/2024

Abstract

A serious and potentially fatal allergic reaction Anaphylaxis requires prompt diagnosis and treatment. Accurate prediction of anaphylaxis is vital for preventing severe outcomes and accompanying medical management. This study investigates the creation and application of a hybrid knowledge-based system architecture to predict anaphylaxis. By combining rule-based systems with machine learning models, the goal is to integrate expert medical knowledge with data-driven insights, resulting in a robust, interpretable, and clinically useful predictive tool. Here the development of a hybrid Knowledge-Based Intelligent System Architecture is discussed. Also, this approach addresses a comprehensive study of the KBS-ML system's capacity against traditional prediction methods, providing healthcare providers with timely and accurate identification of anaphylaxis risk, ultimately improving patient outcomes.

Keywords

Anaphylaxis, Knowledge-Based system, Hybrid Architectures, Intelligent Agent, Machine Learning.

1. Introduction

Anaphylaxis represents a critical condition marked by rapid onset symptoms such as breathing difficulties, hives, and hypotension. Symptoms of anaphylaxis can vary. It often includes difficulty in breathing, swelling of the throat and tongue, rapid heartbeat, low blood-pressure, dizziness, and gastrointestinal distress. Prompt recognition and immediate treatment with epinephrine (adrenaline) are crucial to reverse the reaction and prevent of progression to more severe state. Immediate recognition and treatment are essential to avert severe outcomes, including mortality [1]. Management may also involve allergen-specific immunotherapy to desensitize individuals over time. Traditional diagnostic methods depend heavily on clinical expertise and patient history, which may lead to inconsistent diagnosis and treatment. The unpredictability of anaphylaxis and the complexity of its triggers present significant challenges for healthcare providers. Reliable predictive models are needed to aid in early detection and management, reducing the risk of severe reactions. In this study, the object is to develop an idea of a hybrid knowledge-based system architecture that integrates rule-based and machine-learning approaches [2]. This study also validates the effectiveness of the proposed system in predicting anaphylaxis and ensures the model's predictions are interpretable and actionable for clinical application.

2. Related Work

In today's world Neural networks' ability to understand language nuances aids in grasping diverse sentiments on COVID-19 vaccines has helped develop strategies to address negative perceptions [3] leading to higher acceptance rates and improved pandemic management. Web-based knowledge-driven e-learning tools can greatly shorten the learning time for ER and ICU residents [4] improving understanding of evidence-based medicine, and enhance both patient survival rates and overall outcomes. The majority of architectures examined in the literature prioritize prediction accuracy over explainability and trust worthiness, leading to more intricate embedded approaches. Conversely, parallel (PERML) architectures may be employed to develop a more transparent and easier to explain system [5] for both patients and clinicians. Despite the small dataset size, deep learning methods exhibit promising results in identifying valuable features for cohort selection [6]. In the ML-influenced medical sector, decision support systems rely on an inference mechanism, often utilizing methods like SVM, MLP, and decision trees [7]. In order to address the challenge of facing pressure of efficiency and reduced expenses with quality for healthcare, professionals and researchers are turning more frequently to knowledge-based systems and automation within the medical field [8]. An idea is suggested in this study on active risk identification and analysis [16]. The GAOCNN ML model achieves high accuracy in predicting hospital readmissions and length of stay for diabetic, COVID-19, and ICU patients, crucial for healthcare resource planning and pandemic management [17]. Graph neural networks are specialized deep learning architectures designed to model relationships and interactions in complex data structures, such as social networks, biological networks, and healthcare networks. In the context of predicting anaphylaxis, GNNs can model the intricate relationships between various factors contributing to allergic reactions, including genetic predisposition, environmental exposures, and immune system responses [18]. The predictive models can capture complex interactions between several predictors of anaphylaxis risk by representing patient data as a graph structure and utilizing GNNs. This allows creation of more precise and understandable forecasts. Federated learning is a decentralized machine learning technique that maintains data confidentiality and privacy by distributing model training among several data sources (such as clinics and hospitals). Federated learning can help with overcoming privacy concerns and regulatory barriers in healthcare contexts by enabling collaborative model training on distributed datasets without needing for disclosing sensitive patient information. Predictive models for anaphylaxis risk assessment can use a variety of datasets from various healthcare providers by utilizing federated learning approaches [19]. This improv the robustness and generalization of the model while protecting patient privacy and data security. In order to increase patient confidence, comprehension, and acceptance of machine learning models in healthcare settings, explainable AI techniques seek to offer clear and intelligible explanations for model predictions. The XAI approaches can clarify the reasoning behind model predictions in the context of anaphylaxis prediction, emphasizing the most significant elements influencing a person's risk of allergic responses [20]. Healthcare professionals can obtain practical insights into the underlying mechanisms of anaphylaxis with the incorporation of XAI approaches into the models. This allows for individualized preventive measures and well-informed decision-making as well. Secure multiparty computation (SMC) and homomorphic encryption are two examples of privacy-preserving machine learning algorithms that allow for collaborative model training on sensitive healthcare data while maintaining patient privacy [21]. The application of privacy-preserving ML techniques in context of anaphylaxis prediction can speed up model building and validation while maintaining patient-anonymity by enabling secure data sharing and collaborative research across healthcare facilities. Researchers are able to exploit large-scale datasets from many sources to construct powerful predictive models for anaphylaxis risk assessment while sticking to tight privacy standards and ethical guidelines by applying privacy-preserving ML techniques. Deep learning models known as "generative adversarial networks" are made up of two neural networks, one is the generator and the other is discriminator. These two have been trained in a competitive fashion to produce realistic data samples. GANs can be used in healthcare applications to create artificial data samples that depict various allergic reaction scenarios [22], facilitating the training of models and data augmentation on variety of sample datasets. By leveraging GANs for data augmentation, predictive models for anaphylaxis risk assessment can improve generalization performance and robustness, especially in case of limited data availability or class imbalance. Digital health platforms, such as electronic health records (EHRs), mobile health applications, and telemedicine platforms, offer opportunities for integrating predictive models for anaphylaxis risk assessment into routine clinical workflows. By integrating predictive analytics into digital health platforms, healthcare providers can access real-time risk assessments, personalized recommendations, and decision support tools during patient encounters, enhancing clinical decision-making and preventive care delivery. Additionally, digital health platforms can enable remote monitoring, patient engagement, and data-driven interventions, empowering individuals with allergies to actively manage their condition and minimize the risk of allergic reactions. Nanoparticle implications with machine learning includes the Integrating silver nanoparticles (AgNPs) and AI/ML marks a transformative synergy in food analysis. The advanced data processing feature of it unveils a comprehensive approach, overcoming challenges and redefining standards in food safety monitoring [23].

3. Knowledge-Based Systems (KBS) and ML

An important component of artificial intelligence is Knowledge-Based Systems (KBS), which mimic human decision-making. A knowledge base and an inference engine are the two essential parts of a knowledge base system (KBS). The Expert knowledge, rules, and structured data are all contained within the knowledge base, which acts as a repository for domain-specific information. Meanwhile, using the knowledge stored in the system, the inference engine uses logical reasoning to draw inferences and make judgments. By examining patient-data, medical histories, and symptoms, a KBS can supports medical professionals. They simplify difficult decision-making by offering specialized support for diagnosis, treatment planning, and prognosis evaluation. By the help of combining patient-data and allergen information, KBS improves the prediction of anaphylaxis and gives clinicians quick risk assessments and useful insights. In the field of AI known as machine learning, algorithms acquire knowledge. Machine learning improves anaphylaxis prediction by analyzing patient data and allergen interactions, providing early risk assessment and personalized intervention strategies to mitigate severe allergic reactions effectively.

The successful integrations of KBS and ML offer significant potential across diverse domains. In Algorithmic Trading, hedge funds and financial institutions use KBS to define trading strategies based on market rules and regulations [24]. ML algorithms analyze historical market data to identify patterns and optimize trading decisions in real-time, achieving higher profitability and reduced risk. Areas ruling Financial Risk Assessment, KBS consolidates regulatory guidelines and best practices, while ML algorithms analyze extensive datasets to predict risk factors, improving decision-making for loan approvals, investments, and fraud detection [25]. This enhances accuracy and financial outcomes. The Manufacturing Optimization, KBS utilizes manufacturing expertise and standards, paired with ML's real-time sensor data analysis to detect anomalies and predict equipment failures [26]. This integration boosts operational efficiency, reduces downtime, and enhances product quality, driving cost savings and competitiveness. For Customer Relationship Management (CRM), KBS manages customer profiles and rules, while ML analyzes behavior to personalize interactions and predict churn [27]. This enables personalized recommendations, enhancing customer satisfaction, loyalty, and sales revenue. Supply Chain Management utilizes KBS to offers expertise in logistics to manage inventory rules and constraints (e.g., shelf life, demand forecasts), coupled with ML's analysis of demand sales data, seasonality patterns, and external factors (e.g., weather) to optimize inventory levels, reducing stockouts and excess inventory costs forecasting and market trends [28]. Optimizing inventory levels and reducing lead times improves supply chain visibility, agility, and customer service while minimizing costs. In case of Pathology Diagnosis, the Path-AI uses a combination of KBS, which encodes expert knowledge about pathology rules and classifications, and ML techniques like deep learning to analyze histopathology images [29]. This integration enables more accurate and faster diagnosis of diseases like cancer, improving patient outcomes. In term of Customer Service Chatbots companies integrate KBS to structure dialog flows and customer service guidelines. ML algorithms process natural language input from users [30], learning from interactions to provide contextually appropriate responses and resolve issues efficiently. Smart Grids as Energy Management Utility are used by companies to integrate KBS to model grid operations and energy policies [31]. ML algorithms analyze real-time data from smart meters and weather forecasts to optimize energy distribution, predict demand peaks, and balance renewable energy sources more effectively. Legal Research as a Legal Services Platforms like ROSS Intelligence combine KBS with ML algorithms to navigate legal databases and case law [32]. KBS encodes legal principles and precedents, while ML learns to identify relevant cases and provide insights for lawyers, enhancing legal research efficiency.

3.1. Integration of Knowledge-Based Systems (KBS) and ML

KBS is AI system designed to replicate human decision making by using structured knowledge. They include a knowledge base with domain specific information and an inference engine that applies logical rules to solve problems. Its fundamental parts include knowledge bases, inference engines, user interfaces, explanation facilities, rule-based and expert systems, and knowledge acquisition. Such system is described in figure.1 that illustrates how a text goes through the categorization process. Text data is first processed by a Machine Learning Classifier using a pre-trained model. The output is then passed to a Rule-based Classifier, which refines the classification using a set of predefined rules, ultimately categorizing the text into specific categories. ML models commonly used in healthcare includes Support Vector Machines, Neural Networks, Decision Trees, Random Forests, and Naive Bayes. Challenges faced are: Data Quality and Quantity, Interpretability (Black-box nature of some ML models makes it difficult to understand how predictions are made [9] raising concerns about trust and accountability), regulatory and Ethical Issues, Integration with Clinical Workflow etc. Integrating ML with knowledge-based systems combines the strengths of both approaches: such as Enhanced Decision Support, Improved Accuracy, Dynamic Learning, Reduced Uncertainty etc.

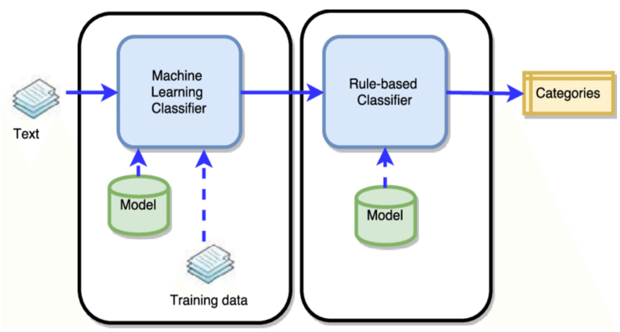

Figure.1: KBS-ML Integration

4. Outline of theory

Before developing the model, there should be a clear idea about knowledgebase system, machine learning and their hybridised approach.

4.1. Development of the Knowledge-Based System

Building the knowledge-based system required close collaboration with medical experts to extract and encode crucial diagnostic rules for predicting anaphylaxis. Utilizing frameworks like Drools and CLIPS, a rule-based system has been developed [11] for architectural improvement. E.g., one rule states, "If a patient has a history of severe reactions to insect stings and presents with swelling and breathing difficulties, the risk of anaphylaxis is high." Another rule specifies, "If a patient with known food allergies exhibits hives and a sudden drop in blood pressure, immediate intervention is necessary." These rules were validated to ensure clinical relevance and reliability. In figure.2 illustration of a clinical decision support system (CDSS) describes fusion of rule-based and case-based reasoning is given. The clinician interacts with the system via a graphical user interface. Rule-based reasoning is involving as a working memory and an inference engine to apply predefined rules. On the other hand, Case-based reasoning includes retrieving, reusing, revising, and retaining past cases to solve new problems. Both reasoning methods access a shared knowledge/case base. Knowledge acquisition, managed by a knowledge engineer, updates this base. The system aims to assist clinicians in making in-formed decisions by leveraging both structured rules and historical case data.

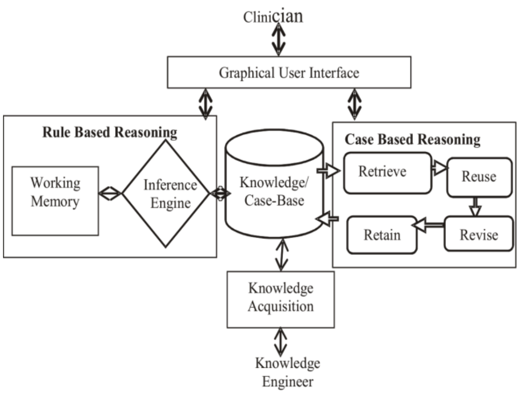

Figure.2: Development of Knowledge-based

4.2. Development of the Machine Learning Model

The development process of Machine Learning model involves selecting algorithms such as Decision Trees, Random Forests, Gradient Boosting Machines, and Neural Networks. Model training included data splitting, cross-validation, and hyperparameter tuning to optimize performance. Figure.3 illustrates about the process of training a machine learning model. Initially, training data consisting of predictor variables (X) and a target variable (Y) is used to create an initial model. This model is then optimized through parameter adjustments to improve its accuracy. After the model is trained, there is possibility to use it for making predictions on new data. The new data, containing new predictor variables (X_new), is fed into the trained model to estimate the target variable (Y_new). This process enables the model to make informed predictions based on the patterns learned from the training data [12]. Thus, analysing feature importance helped identify key predictors, enhancing both interpretability and reliability of the models.

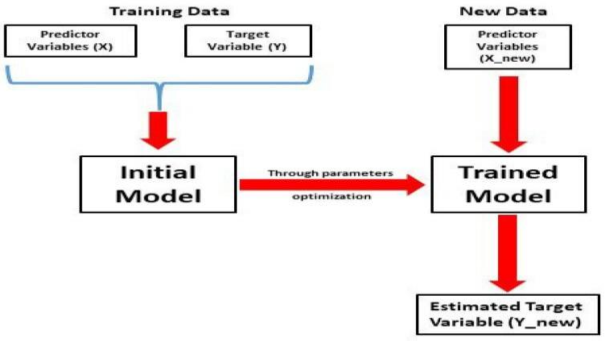

Figure.3: Development of ML model

4.3. Development of Integrated Hybrid Model

This integrated approach ensures robust and accurate predictions, leveraging the strengths of both methodologies to enhance the system’s predictive performance and reliability in clinical applications [5]. The system works as a Clinical Decision System's development and production phases in two sections- System Development & Training and Production System. In the "System Development & Training" phase, the process starts with gathering clinical knowledge, which is expressed in the form of rules (R) like "if condition T1 then suggest action S1." These rules capture expert knowledge and decision-making processes. Alongside this, data is collected, consisting of pairs of inputs (x) and corresponding outputs (y), representing real-world scenarios and outcomes. This clinical knowledge and data are then used to design the AI solution architecture. The AI system is trained using this data, where it learns to map inputs to outputs based on the rules and patterns found in the training data. This training phase involves continuous adjustment to refine the system’s accuracy and reliability. Once the AI solution is well-trained, it moves into the "Production System" phase. Here, the system becomes an operational clinical decision support system, capable of processing new inputs (x) and generating appropriate outputs (y) based on the embedded rules within its architecture.

5. System Design

The system design is discussed in two parts – the Architecture and the Data Flow and Components.

5.1. The Architecture of the Hybrid Knowledge-Based System

In the diagram of figure.4 illustration of the architecture of Hybrid Knowledge-Based System (HKBS) for predictive maintenance and fault diagnosis in equipment is displayed. This system integrates data-driven machine learning models with domain expertise [13]. The Sensors on equipment function by gathering real-time data, such as vibration, temperature, and pressure etc. Then this data is transmitted to central storage system, onboard or in cloud, where preliminary feature extraction can occur. Historical data is used to train predictive models. Important features are extracted from the raw data, and simulated data generation (Design of Experiments) augments the training dataset. Domain knowledge is embedded to create a Health Index, indicating the equipment's health state. The learning algorithm trains models to predict current faults and future trends. The trained model, known as the KML Model, incorporates both data-driven insights and domain knowledge. The Health Index evaluates the equipment's current state. Real-time data is continuously fed into the system, where features are extracted, and the KML model predicts the current state and remaining useful life (RUL) of the equipment. These predictions are visualized for end-users. The system's performance is regularly evaluated. If performance drops below a threshold, it triggers model retraining [14]. Continuous feedback updates domain knowledge and improves the Health Index, ensuring high accuracy and reliability in predictive maintenance.

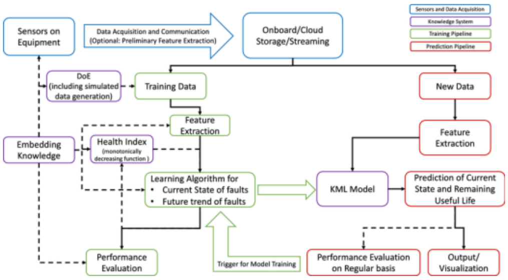

Figure.4: Hybrid (KML) Architecture

5.2. Data Flow and Components

The second important part of system design is about introducing Data Flow and Component. The flow of information occurs along with the function on each component as followed. Problem Definition**: -** Function f:

x → y defines the relationship or function that maps input x to output y, Data: - Input-output pairs (x1, y1), ..,(xn, yn) represent the dataset utilized in testing and training the ML model, Knowledge: - Rules in form R1: if T1 then S1,…, Rn: if Tn then Sn encodes domain-specific rules and knowledge used in the system, ML Solution Building: - Involves rule embedding, training, and validation of the ML model. Integration of rules within the ML model during the solution-building phase, Solution: - The final system that takes inputs x, processes them using the integrated rule-based and ML approaches, and provides outputs y and at the end Monitoring & Adjustment: - Post-deployment monitoring and adjustment of the system to ensure performance and reliability over time.

6. Methodology

The methodology consists of various segmentations- dataset, feature selection and model development, target, finding objective and the feature extraction.

6.1. Dataset

Datasets comprising patient demographics, medical history, allergic sensitivities, environmental factors, and previous anaphylactic events are collected from electronic health records (EHRs), wearable devices, and patient reported outcomes. Data preprocessing techniques such as normalization, imputation, and feature engineering are employed to handle missing values, standardize data formats, and extract relevant features for model training. Finding suitable datasets for predicting anaphylaxis and implementing preventive measures with machine learning can be challenging due to the sensitive nature of healthcare data and the relatively rare occurrence of anaphylactic events. Healthcare institutions may provide access to de-identified EHR datasets for research purposes. These datasets typically contain patient demographics, medical history, diagnoses, medications, laboratory results, and encounter information, which can be used to identify individuals with a history of anaphylaxis and extract relevant features for predictive modeling. Government agencies and public health organizations may offer publicly available datasets related to allergic reactions, emergency department visits, hospital admissions, mortality records, and population health surveys. Examples consists of Centers for Disease Control and Prevention (CDC), National Institutes of Health (NIH), and World Health Organization (WHO). Academic institutions, research consortia, and pharmaceutical companies conduct studies and clinical trials focused on allergic diseases, including anaphylaxis. These studies may provide access to anonymized datasets with detailed patient information, intervention outcomes, and adverse event reports. Allergy and immunology societies often maintain patient registries or databases to track individuals with specific allergic conditions, including anaphylaxis. The research collaborations that enable collaboration with these organizations or grant access to their databases can provide insightful information for predictive modeling and preventive measures too. Rich datasets are appropriate for predictive analytics which maybe produced by wearable devices with sensors tracking the physiological parameters (e.g., heart rate, respiration rate, skin conductance) and by mobile applications to track symptoms and allergen exposure. Working with app developers or device makers could make it easier to obtain these datasets. The Healthcare Cost and Utilization Project (HCUP), Observational Medical Outcomes Partnership (OMOP), and Medical Information Mart for Intensive Care (MIMIC) are few examples of online platforms and repositories devoted to healthcare data sharing and collaboration, may contain research datasets related to anaphylaxis [33]. These platforms frequently offer carefully chosen datasets together with ethical standards and suitable data use agreements. Researchers may use data modeling techniques or create synthetic datasets to simulate the features of anaphylaxis and its related hazards in situations where the real-world data sources are few or unavailable owing to privacy issues. For algorithm creation and validation, synthetic datasets can be helpful, even though they are not as representative as real-world data.

6.1.1 Caution Handling

It is crucial to follow data privacy laws (such as the Health Insurance Portability and Accountability Act and the General Data Protection Regulation) and secure the required approvals from institutional review boards (IRBs) or ethics committees before accessing and using healthcare datasets for research purposes. Furthermore, it is also required to make sure that the appropriate data anonymization and de-identification methods are used in order to protect patient-confidentiality and comply with ethical-standards [34]. Sometimes it can be more difficult to get real-time datasets for machine learning preventive measures and anaphylaxis prediction because of reasons such as the Tight privacy laws like the General Data Protection Regulation (GDPR) in the European Union and the Health Insurance Portability and Accountability Act (HIPAA) in the United States control the handling of healthcare data, especially real-time patient information. Working together with healthcare organizations, data custodians, and regulatory agencies is necessary to access real-time datasets while guaranteeing compliance with these requirements. The availability of real-time healthcare data, such as electronic health records (EHRs), for research purposes may be limited by institutional restrictions, data ownership disputes, and patient privacy and consent concerns. It is possible that healthcare organizations will not readily share real-time data with systems outside of their own.

6.1.2 Potency

Notwithstanding these obstacles, real-time datasets for anaphylaxis prediction and prevention studies may be accessed through the following channels such as Research possibilities using anonymized datasets may arise from collaborating with hospitals, clinics, and healthcare systems that gather real-time patient data via electronic health records and monitoring systems. Facilitating data access and exchange can be achieved through collaborating with healthcare providers and gaining the necessary approvals. Participating in ongoing clinical trials or observational studies focused on allergic diseases and anaphylaxis may provide access to real-time patient data collected during the study period. Collaborating with principal investigators and research teams involved in these studies can facilitate data sharing and collaboration. Collaborating with wearable device manufacturers or mobile health app developers that collect real-time physiological data relevant to anaphylaxis (e.g., heart rate, respiratory rate, skin conductance) may provide access to real-time datasets for research purposes. Exploring partnerships with companies in the digital health space can yield opportunities for accessing real- time data streams. Leveraging health data aggregation platforms and networks that facilitate real-time data sharing and collaboration among healthcare institutions, research organizations, and technology partners may provide access to aggregated real-time datasets for research purposes. Examples include health information exchanges (HIEs), research consortia, and data-sharing initiatives focused on allergic diseases. While obtaining real-time datasets for research on anaphylaxis prediction and prevention may be challenging, proactive engagement with stakeholders, collaboration with healthcare partners, and adherence to ethical and regulatory requirements can facilitate access to valuable data sources for advancing scientific knowledge and improving patient outcomes. Database suggestion includes ESCMID, JADER, CDC, NIH, WHO, PubMed Central etc.

6.2. Feature Selection and Model Development

Feature selection methods such as recursive feature elimination (RFE), principal component analysis (PCA), and correlation analysis are utilized to identify the most informative predictors of anaphylaxis risk. Various ML algorithms including logistic regression, support vector machines (SVM), random forests, and gradient boosting machines (GBM) are trained on the preprocessed data to build predictive models. Model performance is evaluated using metrics such as accuracy, precision, recall, and area under the receiver operating characteristic curve (AUC-ROC). Important steps in the study process include feature selection, model construction, and machine learning in the context of anticipating anaphylaxis and putting preventive measures in place, described by the following.

6.2.1. Feature Selection

The process of selecting features entails determining which features, out of the provided dataset, are most informative in predicting the occurrence or severity of anaphylaxis. Careful feature selection is necessary to create prediction models that are accurate and easy to understand, given the complexity of allergic reactions and the variety of factors that can increase the risk of anaphylaxis. This is an example of how feature selection can be done as the following.

i. Domain Knowledge:

To find potential predictors linked to anaphylaxis risk, allergy and immunology domain specialists are consulted first. These could include biological characteristics (heart rate, respiratory rate), demographics (age, sex), medical history (prior allergic reactions, comorbidities), allergen exposure (food allergies, insect stings), environmental triggers (pollen, pollutants), and medication use (allergy medications, immunotherapy).

ii. Exploratory Data Analysis (EDA):

Analyzing exploratory data to find out how the desired variable (such as the severity of anaphylactic reactions) and the available attributes are related. Finding patterns, correlations, and outliers that could influence feature selection is aided by this.

iii. Methods of Feature Importance:

To rank the predictors according to their significance to the target variable, feature importance approaches such univariate statistical tests, correlation analysis, information gain, or tree-based feature importance are used. This aids in setting the features' relative priorities for the prediction models.

iv. Reduction of Dimensionality:

To cut down on features without losing the most important data, consider dimensionality reduction techniques like principal component analysis (PCA) or feature clustering. This can facilitate model interpretation and make model training easier.

v. Methods for Regularization:

In the training phase of a model, apply regularization techniques like L1 (Lasso) or L2 (Ridge) regularization to penalize irrelevant or redundant information [35] and promote the model coefficients' sparsity. This improves model generalization performance and helps avoid overfitting.

vi. Recursive Feature Elimination (RFE):

In order to repeatedly remove less informative features from the dataset based on model performance indicators (e.g., accuracy, AUC-ROC), implementation of recursive feature reduction techniques is required. This iterative process helps identify the subset of features that contribute most significantly to the predictive performance of the models. By carefully selecting informative features and reducing the dimensionality of the dataset, researchers can build more robust and interpretable predictive models for anaphylaxis risk assessment and prevention.

6.2.2. Model Development

Once the relevant features have been identified, the next step is to develop predictive models that can effectively leverage these features to predict anaphylaxis occurrence or severity of it. The approach for model development includes parameters such as selection, training, evaluation, interpretability, ensembleing methods and cross validation which is discussed in the following.

i. Model Selection:

Choosing an appropriate machine learning algorithm based on the nature of the problem (e.g., binary classification for predicting anaphylaxis occurrence, regression for predicting severity scores) plays a very important role. Some commonly used algorithms include logistic regression, support vector machines (SVM), random forests, gradient boosting ma-chines (GBM), and deep learning models.

ii. Model Training:

The very next step is splitting the dataset into training and validation sets to train and evaluate the performance of the models. Applying the appropriate preprocessing techniques such as feature scaling, handling missing values, and encoding categorical variables are performed as per the need. Next the selected models are trained using the training data and tune hyperparameters using techniques like grid search or randomized search to optimize model performance etc.

iii. Model Evaluation:

Evaluate the trained models using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve, or mean squared error (MSE) depending on the task (classification or regression). Compare the performance of different models and select the best-performing model(s) for deployment.

iv. Model Interpretability:

Ensuring selected models are interpretable and provide insights into the factors contributing to anaphylaxis risk. Techniques such as feature importance analysis, partial dependence plots, and SHap-ley Additive explanations values help interpret the model predictions and understand the impact of in-dividual features on the outcomes.

v. Ensemble Methods:

It is advisable to employ ensemble learning methods like stacking, boosting, or bagging to integrate various base models and enhance prediction abilities. In order to reduce model bias and variation and produce more accurate and dependable predictions, ensemble approaches might be used.

vi. Cross-Validation:

To evaluate the models' generalization abilities and make sure they are not overfitting to the training set, do cross-validation. For unknown data, methods like stratified cross-validation and ask-fold cross-validation can yield more precise estimates of the model's performance.

6.3. Target

The target variable can be binary, meaning that it indicates if a person has anaphylaxis (1), does not have anaphylaxis (0), or is multiclass if various anaphylactic severity levels are taken into account. Other pertinent clinical outcomes, the time of occurrence, the severity of the symptoms, and the response to therapy could all be included in the target variable. Healthcare professionals can reduce the incidence and severity of allergic reactions and improve patient outcomes by identifying people who are at high risk of anaphylaxis and implementing targeted preventive interventions, such as allergen avoidance strategies, prescribing epinephrine auto-injectors, patient education, and emergency action plans. The target variable can be thought of as a scale to measure the severity of anaphylactic reactions if it is expressed numerically:

i. Severity Scale:

The target variable in this representation has a number value assigned to it that represents the individual's level of allergic response severity. By providing a numerical value to represent the intensity of the allergic reaction, this scale makes it possible quantification of the severity of anaphylactic reactions and improve the prediction and management of allergic occurrences:

Table 1. Table for scaling the severity

Sl.No.	Value	Indication	Alert	Symptom Details
i.	0	No reaction	Null	Null
ii.	1	Mild reaction	Symptoms are localized and relatively minor	Localized hives, itching, or mild swelling
iii.	2	Moderate reaction	Symptoms are more pronounced	respiratory distress, Gastrointestinal symptoms, extensive swelling

iv.	3	Severe reaction	Requiring immediate medical intervention	Severe respiratory distress, cardiac arrest, unconsciousness, other life-threatening symptoms

ii. Reaction Score:

Calculating a reaction score based on the intensity and duration of symptoms experienced by the individual. The severity rating systems or clinical evaluation instruments used in allergy and immunology research may be taken as the source of this score.

iii. Treatment Response Index:

Putting a number on how well a patient's symptoms are reduced following medical intervention in order to quantify their response to the treatment.

Table 2. Table for scaling responses

Sl.No.	Value	Indication
i.	0	No improvement
ii.	1	Partial improvement
iii.	2	Complete resolution of symptoms

iv. Risk Probability:

Analyzing someone's chance of developing anaphylaxis in a specific amount of time and expressing the result as a probability or risk score. Predictive models trained on different risk factors and predictors linked to anaphylactic reactions may yield this. If there was a predictive algorithm able enough to tell likely hood of someone getting anaphylaxis in the coming twenty-four hours. This model is trained using a variety of anaphylactic reaction risk indicators and predictions.

Probability Representation: The model gives outputs on probability score between 0 and 1.

Where it may be

Fraction: Ranges from 0 to 1 or
Decimal: Ranges from 0 to 1 or
Percentage: Ranges from 0% to 100%

Table 3. Table for Probability Scaling

Sl.No.	Value	Indication	**Suggestion**
i.	0	Indicates very low likelihood	No risk of anaphylaxis within the timeframe
ii.	1	Indicates very high likelihood	High risk of anaphylaxis within the timeframe

The Model may give output as a probability score of 0.75, indicating a 75% likelihood of experiencing anaphylaxis within coming 24 hours.

Risk Score Representation: The model assigns a risk score on the basis of estimated probability. Scaling score can have a specific range, such as 0 to 100, for ease in interpretation.

Threshold: It is established to categorize individuals into different risk levels based on their probability or risk score.

Table 4. Table for Threshold measurement

Sl.No.	Risk Intensity	Risk or Probability Score
i.	Low	< 0.25
ii.	Moderate	0.25 ≤ and < 0.75
iii.	High	≥ 0.75

Interpretation:

People with higher probability or risk ratings are thought to be more likely to develop anaphylaxis within the allotted period. Medical providers can use these scores to rank actions for higher-risk patients, such as prescribing preventative medicine or suggesting lifestyle changes. For example, if a person receives a probability score of 0.75 or a risk score of 75 from the model, it indicates that they have a high chance of developing anaphylaxis in the next 24 hours, calling for prompt medical care or preventive measures.

v. Resource Utilization Index:

Analyzing the quantity of medical resources used to treat anaphylactic instances, such as the quantity of ER visits, hospital stays, or necessary medical procedures, and reporting the results as scores or numerical counts. Each of these numerical representations provides a quantitative measure of the target variable, enabling more detailed analysis and prediction of anaphylactic reactions and their associated outcomes.

6.4. Finding Objective

By following the mentioned steps, this paper aims to systematically address the research objectives, generate actionable insights, and contribute to the advancement of knowledge in anaphylaxis prediction and prevention using machine learning approaches.

i. Problem Definition and Research Objectives:

Clearly define the research problem: Predict anaphylaxis occurrence and severity using machine learning to im plement timely preventive measures. Specify research objectives: Develop predictive models to accurately iden tify individuals at risk of anaphylaxis, effectiveness of preventive interventions, and improve patient outcomes.

ii. Data Collection and Preprocessing:

Obtain datasets from multiple sources, including electronic health records (EHRs), public health databases, wearable devices, and clinical trials. Clean and preprocess the data to address missing values, outliers, and inconsistencies. Perform feature engineering to extract relevant features such as demographic information, medical history, allergen exposure, physiological parameters, and environmental factors.

iii. Feature Selection:

Apply feature selection techniques such as univariate analysis, correlation analysis, and recursive feature elimination (RFE – a technique used in machine learning by selecting the most relevant features and reduces the potential for overfitting. It is often used with algorithms such as linear regression, support vector machines, and decision trees etc. Although it is Computationally Intensive which suggests that it can be computationally expensive, especially with large datasets or many features. Another problem of it is being Model Dependent which suggests that selected features may depend on the initial model used for ranking.) is an effective way to identify the most informative predictors of anaphylaxis risk. Consider domain knowledge and expert input to prioritize features relevant to allergic reactions and anaphylaxis.

iv. Suggesting Model Development:

Table 5. Table for Steps for Model Development

Sl.No.	Steps	Suggestion
i.	Select appropriate machine learning algorithms based on the research objectives and dataset characteristics	Logistic regression, decision trees, ensemble methods
ii.	Split the dataset into training, validation, and test sets for model development, evaluation, and validation.	set size, record size
iii.	Train the selected models using the training data and optimize hyperparameters using techniques.	grid search, Bayesian optimization
iv.	Evaluate model performance using relevant metrics and compare multiple models to identify the best-performing one(s).	Relevant metrics e.g.: - accuracy, precision, recall, F1- score, AUC-ROC)

v. Interpretation and Validation:

Interpret model predictions to understand the relative importance of features in predicting anaphylaxis risk. Validate the predictive models using cross-validation techniques to assess their generalization performance on unseen data. Ensure model interpretability and provide actionable insights for healthcare providers to inform preventive interventions.

vi. Implementation and Deployment:

Implement the predictive models in real-world healthcare settings, integrating them into clinical decision support system or electronic health record systems. Collaborate with healthcare stakeholders to deploy preventive interventions based on model predictions, such as allergen avoidance strategies, patient education, and emergency action plans. To guarantee a successful integration into clinical practice, implementation process must be under observation and offer continuous support.

vii. Evaluation and Impact Assessment:

Analyze how well prediction models and preventive measures work to lower the frequency and severity of anaphylactic reactions. Evaluate results such patient adherence to preventative treatment, the use of healthcare resources, and modifications to clinical practice recommendations. Analyze how the study's conclusions affect patient treatment, medical results, and public health programs, how the conclusions affect patient treatment, medical results, and public health programs.

6.5. Feature Extraction

Feature extraction consists of elements such as Data Understanding, Demographics, Feature Selection, Validation, Exposure, History, Environment etc. The following is a brief discussion of these parametric.

i. Data Understanding:

This entails on the basis of learning every information about the accessible data sources such as electronic health records (EHRs [10]), databases pertaining to clinical trials, wearable technology, and public health. Also, to determine the kinds of information that are present in each dataset, including physiological measurements, medical history, allergen exposure, environmental factors, and clinical results pertaining to anaphylaxis.

ii. Demographic Features:

Take demographic data from patient records, including age, sex, ethnicity, and socioeconomic status. Use either ordinal or one-hot encoding to encode categorical variables (such as sex or ethnicity). To make sure numerical variables are on a similar scale, normalize them (e.g., age).

iii. Feature Selection:

To find possible significant factors linked to the risk and severity of ana-phylaxis, domain knowledge and expert advice are used. Examine the dataset's distributions, relationships, and trends using exploratory data analysis (EDA- it guarantees further analyses are based on a thorough understanding of data by transforming raw data into actionable insights, affecting the outcome of initiatives driven by data) to find potential traits for extraction.

iv. Feature Validation:

To guarantee the relevance and predictive capacity of the chosen features, validate them using statistical testing, correlation analysis, or domain expert assessment. Utilize feature selection algorithms (such RFE and L1 regularization) to determine which subset of features minimizes overfitting and makes the greatest contribution to the prediction of anaphylaxis. To improve predictive power, combine or alter current features to create new ones. Determine derived parameters like body mass index (BMI) using weight and height data, or allergen exposure intensity scores based on allergen concentrations and time of exposure. Use polynomial features or interaction terms to capture nonlinear correlations between variables and the risk of anaphylaxis.

v. Feature Engineering:

Improves the predictive strength and relevance of current features to the prediction of anaphylaxis by creating new ones or transforming old ones. Techniques for feature engineering consist of developing binary markers for comorbid diseases or allergen exposures (e.g., food allergies, asthma), combining time-related data (such as the frequency of prior allergy responses and seasonal changes in allergen exposure) and utilizing natural language processing (NLP- it is a sub field of AI) to extract pertinent data from unstructured data sources like clinical notes or free-text entries.

vi. Allergen Exposure:

Based on the findings of allergy tests or the allergies that the patient has indicated, is listed as known allergens. Make binary indicators for allergen exposures that show whether a person is known to be allergic to a particular allergen (such as shellfish or peanuts). Provide temporal details about recent exposure events, like when and how often you ingested allergens or were stung by insects.

vii. History regarding Allergen Exposure:

Record the medical background of each person, encompassing past allergic reactions, allergies conditions diagnosis (e.g., asthma, eczema), and medication use (e.g., immunotherapy, allergy drugs). Keep track of your history of allergy exposure, including known allergies, past reactions to certain allergens, and recent exposure incidents (such as eating intake or insect stings).

viii. Physiological Parameters:

Take physiological parameters, such as blood pressure, skin temperature, heart rate, respiration rate, and oxygen saturation, that are related to the risk of anaphylaxis. Integrate data from medical sensors or wearable devices to record physiological indicators in real-time, allowing for ongoing monitoring of people who may experience allergic reactions. To identify temporal patterns and trends, compute time-series related characteristics (such as rate of change) or summary statistics (such as mean, median, mode, Interquartile Range and standard deviation) from physiological data.

ix. Environmental Factors:

Features relate to environmental exposures that could trigger allergic reactions, such as pollen counts, air pollution levels, temperature, humidity, geographic location, and atmospheric pressure. Incorporate and Integrate data from environmental monitoring stations or public health databases to capture spatial and temporal variations in environmental exposures. Calculate environmental indices or aggregates (e.g., pollen season severity index) to summarize impact of multiple environmental factors on anaphylaxis risk.

7. Data Pre-processing:

Encode categorical features using one-hot encoding or label encoding to represent categorical variables as numerical values suitable for modelling. Pre-process data to handle missing values, outliers, and data quality issues using ap-propriate techniques such as imputation, outlier detection, and data cleaning. Validate selected features through statistical analysis, correlation analysis, or domain expert review to ensure their relevance and predictive power. Use feature importance techniques (e.g., random forest feature importance, permutation importance) to assess the contribution of each feature to the predictive models and identify the most informative predictors of anaphylaxis risk. Standardize numerical features by scaling them to a common range (e.g., mean normalization, min-max scaling). Utilize imputation methods to handle missing values, like mean imputation, median imputation, or predictive imputation based on correlated features. By performing feature extraction systematically and comprehensively, identifying is possible in terms of informative predictors of anaphylaxis risk and severity, which can be used to develop accurate and interpretable predictive models for preventive interventions.

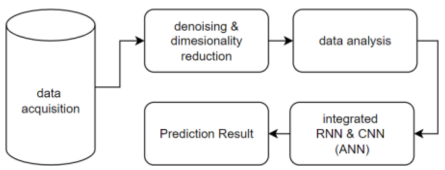

Figure 2:-Data preprocessing

8. Data Collection and Preparation

The system relies on diverse data sources such as Electronic Health Records (EHR), patient histories, demographic data, and genetic information. These sources provide a comprehensive dataset for analysis. Data pre-processing involves cleaning to eliminate inconsistencies, normalization to standardize formats, and feature engineering [10] to generate relevant variables for model training.

9. Implementation Details

The system utilizes technologies like Drools for rule-based logic and libraries like TensorFlow and Scikit-learn for machine learning. Developed in a Python-based environment, the system is designed for seamless deployment in clinical settings. The user interface, tailored for healthcare providers, enables easy data input and provides clear, actionable predictions, ensuring practical and efficient clinical use.

10. Validation

To assess the performance on the basis of validation of the hybrid knowledge-based system, several such evaluation metrics that are utilized are mentioned below.

10.1. Accuracy

This measures the overall correctness of the predictions by calculating the ratio of correctly predicted instances to the total instances following the equation (i).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{(i)}$$

10.2. Precision (Positive Predictive Value)

It evaluates the accuracy of the positive predictions by determining the ratio of true positive instances to the sum of true and false positive instances following the equation (ii).

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{(ii)}$$

10.3. Recall (Sensitivity or True Positive Rate)

Recall measures the system's ability to correctly identify positive instances by calculating the ratio of true positive instances to the sum of true positives and false negatives following the equation (iii).

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{(iii)}$$

Where, TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

10.4. F1-Score

It provides a balance between precision and recall, calculated as the harmonic mean of precision and recall ranging from 1 (best value = perfect precision and recall) to 0 (worst) following the equation (iv).

$$\text{F1} - \text{Score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \qquad \text{(iv)}$$

10.5. ROC-AUC (Receiver Operating Characteristic - Area Under the Curve)

The ROC curve plots the Recall against the False Positive Rate (1 - Specificity) at various threshold settings. The area under this curve (ROC-AUC: stands for Receiver Operating Characteristic - Area Under the Curve. It may help with performance of binary classification model. It helps to evaluate how well the model separates the positive and negative classes regardless of the chosen threshold) indicates the model's ability to discriminate between positive and negative instances, with higher values indicating better performance [15].

11. Results

The results revealed the performance matrices and comparative analysis showcased that while rule-based systems exhibited high interpretability and accuracy in capturing explicit medical rules, they sometimes lacked the complexity to handle nuanced patterns in the data. On the other hand, Machine Learning models demonstrated impressive predictive capabilities, although often lacked interpretability, hindering their clinical application. The hybrid system, however, emerged as a promising solution with respect to others by combining the strengths of both approaches identifying at-risk patients, initiate timely interventions, and optimize treatment strategies. Overall, the results underscored the potential of the hybrid knowledge-based system as a valuable tool in enhancing anaphylaxis prediction and improving patient outcomes along with healthcare provider utilizing the KBS. The graph in figure.5 compares the accuracy of three models relative to data volume: Knowledge-based model maintains constant accuracy, Knowledge-embedded Machine Learning model increases accuracy

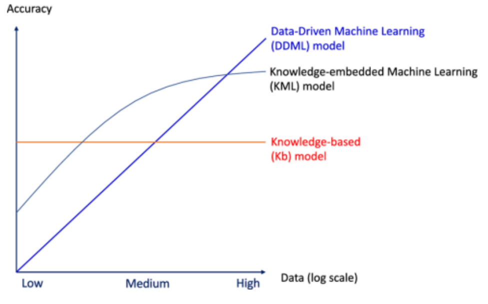

Figure.5: Comparative Analysis

with more data but plateaus, and Data-Driven Machine Learning model continuously improves accuracy as data volume increases.

12. Conclusion

There are intriguing potential uses for the hybrid knowledge-based system (HKBS) in clinical decision-making and anaphylaxis prediction. In order to improve patient outcomes, the HKBS provides a clear framework for identifying at-risk patients and directing prompt measures which is accomplished by combining rule-based and machine learning approaches. Doctors can understand the reasoning behind forecasts due to its interpretability, which encourages decision-making confidence and permits tailored interventions. However, resolving data-related difficulties, like data quantity and quality, is necessary to improve the HKBS's reliability and generalizability across different patient groups. It must also be demonstrated to be successful in a range of clinical settings and have a seamless integration into existing workflows in order to gain widespread acceptance. If the HKBS is consistently observed and updated, it will continue to be applicable and efficient over time.

13. Future Scope

The suggested hybrid model offers a fresh take on anaphylaxis prediction and signifies a paradigm shift in healthcare. It blends the predictability of data-driven models with the interpretability of rules by fusing rule-based systems with machine learning methods. This combination makes it possible to assess risks in a nuanced way, which facilitates prompt interventions to avert severe allergic reactions and greatly improve patient outcomes. The hybrid model has enormous potential to improve patient safety and healthcare delivery because of its capacity to interpret intricate patterns in anaphylactic reactions. The work also provides intriguing new directions for future investigation. Investigating cutting-edge machine learning methods like reinforcement learning and deep learning could uncover hidden risk variables and improve prediction accuracy. Furthermore, examining the hybrid model's scalability and practical use in clinical settings provides insightful information on its efficacy and usability in real-world scenarios. The creation and implementation of user-friendly software platforms could be sped up by large-scale trials and collaborations with industry players, allowing for a smooth transition into standard clinical practice. All things considered, the hybrid model that has been suggested is a shining example of innovation in anaphylaxis prediction, with the potential to significantly alter patient treatment and healthcare decision-making.

Acknowledgements

We would like to express our deepest gratitude to the technical resources of Department of Information Technology, Kalyani Government Engineering College, Kalyani, Nadia - 741235, West Bengal, India and Department of Computer Science and Engineering, College of Engineering and Management Kolaghat, KTPP Township, Purba Medinipur - 721171, West Bengal, India.

References

1. Jawad, B.N., Shaker, S.M., Altintas, I., Eugen-Olsen, J., Nehlin, J.O., Andersen, O. and Kalle-mose, T. Development and validation of prognostic machine learning models for short-and long-term mortality among acutely admitted patients based on blood tests. Scientific Reports, 14(1), p.5942, 2024.
2. Birgit Obermeier, “Hybrid AI- Combined use of knowledge and data”, Plattform Lernende Systeme – Germany's Platform for Artificial Intelligence, Series: AI at a Glance, National Academy of Science and Engineering, 30 November 2023.
3. Saxena, R.R., 2024. Examining Reactions about COVID-19 Vaccines: A Systematic Review of Studies Utilizing Deep Learning for Sentiment Analysis. Authorea Preprints.
4. Riano, D., Real, F. and Alonso, J.R. “Improving resident's skills in the management of circulatory shock with a knowledge-based e-learning tool.”, International journal of medical informatics, 113, pp.49-55, 2018.
5. Kierner, S., Kucharski, J. and Kierner, Z. “Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review.”, Journal of Bio-medical Informatics, p.104428, 2023.

6. Segura-Bedmar, I. and Raez, P. “Cohort selection for clinical trials using deep learning models.”, Journal of the American Medical Informatics Association, 26(11), pp.1181-1188, 2019.

7. Bal, M., Amasyali, M.F., Sever, H., Kose, G. and Demirhan, A. “Performance evaluation of the machine learning algorithms used in inference mechanism of a medical decision support system.”, The Scientific World Journal, 2014.

8. Srivastava, S.R. and Dubey, S. Knowledge-based systems in medical applications. In Computa-tional intelligence and its applications in healthcare (pp. 189-215). Academic Press, 2020

9. Pulicharla, M.R., “Explainable AI in the Context of Data Engineering: Unveiling the Black Box in the Pipeline.”, International Journal of Innovative Science and Research Technology, Volume 9, ISSN No: - 2456-2165, 2024.

10. Amirahmadi, A., Ohlsson, M. and Etminani, K. “Deep learning prediction models based on EHR trajectories: A systematic review.”, Journal of biomedical informatics, p.104430, 2023.

11. Sivasankari, S., Punnoose, D. and Krishnamoorthy, D. “A comparative study on the performance of rule engines in automated ontology learning: a case study with erythemato-squamous disease (ESD).”, International Journal of Intelligent Unmanned Systems, 8(4), pp.267-280, 2020.

12. Habibi, L.N., Matsui, T. and Tanaka, T.S. “Critical evaluation of the effects of a cross-validation strategy and machine learning optimization on the prediction accuracy and transferability of a soybean yield prediction model using UAV-based remote sensing.”, Journal of Agriculture and Food Research, 16, p.101096, 2024.

13. Farbiz, Farzam, "Knowledge-embedded machine learning and its applications in smart manufac-turing.", Journal of Intelligent Manufacturing 34.7 (2023): 2889-2906, 2023.

14. Kierner, S., Kucharski, J. and Kierner, Z., “Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review.”, Journal of Bio-medical Informatics, p.104428, 2023.

15. Mustaqeem, M., Mustajab, S. and Alam, M. “A hybrid approach for optimizing software defect prediction using a gray wolf optimization and multilayer perceptron.”, International Journal of Intelligent Computing and Cybernetics, 2024.

16. Carrell, D.S., Gruber, S., Floyd, J.S., Bann, M., Cushing-Haugen, K., Johnson, R., Graham, V., Cronkite, D., Hazlehurst, B., Felcher, A.H. and Bejin, C.A., 2021. Improving methods of identifying anaphylaxis for medical product safety surveillance using natural language processing and machine learning. Pharmacoepidemiology and Drug Safety, 30, pp.16-17.

17. Tavakolian, A., Rezaee, A., Hajati, F. and Uddin, S., 2023. Hospital Readmission and Length-of-Stay Prediction Using an Optimized Hybrid Deep Model. Future Internet, 15(9), p.304.

18. Bedraoui, A., Suntravat, M., El Mejjad, S., Enezari, S., Oukkache, N., Sanchez, E.E., Galan, J.A., El Fatimy, R. and Daouda, T., Medicine in Drug Discovery.

19. Vats, S., Kukreja, V. and Mehta, S., 2024, March. A New Era in AgriTech: Federated Learning CNN for Jute Leaf Disease Identification. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) (Vol. 2, pp. 1-6). IEEE.

20. Gupta, J. and Seeja, K.R., 2024. A Comparative Study and Systematic Analysis of XAI Models and their Applications in Healthcare. Archives of Computational Methods in Engineering, pp.1-26.

21. Ahammed, M.F. and Labu, M.R., 2024. Privacy-Preserving Data Sharing in Healthcare: Advances in Secure Multiparty Computation. Journal of Medical and Health Studies, 5(2), pp.37-47.

22. Lange, L., Wenzlitschke, N. and Rahm, E., 2024. Generating Synthetic Health Sensor Data for Privacy-Preserving Wearable Stress Detection. arXiv preprint arXiv:2401.13327.

23. Moulahoum, H. and Ghorbanizamani, F., 2024. Navigating the development of silver nanoparticles based food analysis through the power of artificial intelligence. Food Chemistry, p.138800.

24. Borch, C., 2022. Machine learning, knowledge risk, and principal-agent problems in automated trading. Technology in society, 68, p.101852.

25. Setiono, R., Mues, C. and Baesens, B., 2006. Risk management and regulatory compliance: A data mining framework based on neural network rule extraction. ICIS 2006 Proceedings, p.7.

26. Kumar, R., Sangwan, K.S., Herrmann, C. and Ghosh, R., 2023. Development of a cyber physical production system framework for smart tool health management. Journal of Intelligent Manufacturing, pp.1-30.

27. Mazumdar, B.D. and Roy, S., 2021. Multi-Agent Paradigm for B2C E-Commerce. In Artificial Intelligence and Machine Learning in Business Management (pp. 29-52). CRC Press.

28. Kharfan, M., Chan, V.W.K. and Firdolas Efendigil, T., 2021. A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. Annals of Operations Research, 303(1), pp.159-174.

29. McNeil, C., Wong, P.F., Sridhar, N., Wang, Y., Santori, C., Wu, C.H., Homyk, A., Gutierrez, M., Behrooz, A., Tiniakos, D. and Burt, A.D., 2024. An End-to-End Platform for Digital Pathology Using Hyperspectral Autofluorescence Microscopy and Deep Learning-Based Virtual Histology. Modern Pathology, 37(2), p.100377.

30. Borek, C., 2024. Comparative evaluation of llm-based approaches to chatbot creation.

31. Khalid, M., 2024. Smart grids and renewable energy systems: Perspectives and grid integration challenges. Energy Strategy Reviews, 51, p.101299.

32. Morison, J. and Mclnerney, T., 2024. When should a computer decide? Judicial decision-making in the age of automation, algorithms and generative artificial intelligence. Research Handbook on Judging and the Judiciary (Elgar-Routledge, La

33. Ostropolets, A., 2023. Generating Reliable and Responsive Observational Evidence: Reducing Pre-analysis Bias. Columbia University.

34. Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A. and Lovis, C., 2019. Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. Journal of medical Internet research, 21(5), p.e13484.

35. Rahmat, F., Zulkafli, Z., Ishak, A.J., Abdul Rahman, R.Z., Stercke, S.D., Buytaert, W., Tahir, W., Ab Rahman, J., Ibrahim, S. and Ismail, M., 2024. Supervised feature selection using principal component analysis. Knowledge and Information Systems, 66(3), pp.1955-1995.