P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

Explainable Generative AI for Real-Time HCC Risk Scoring from FHIR Streams

Daniel Thompson
Asst Professor

Abstract

Early detection and timely treatment of hepatocellular carcinoma (HCC) are critical to improving patient outcomes, yet existing risk-scoring models are poorly suited to clinical practice: their prediction horizons are too long, their input data are often unavailable at the point of care, and their outputs lack real-time interpretability. This study introduces a predictive model that computes HCC risk scores in real time, allowing seamless integration into routine clinical decision-making workflows. To address the interpretability limitations inherent to deep learning, we apply an Explainable Generative Modeling Framework — a novel combination of generative modeling and explainability techniques — to reveal the mechanisms underlying the model's predictions. Performance evaluation, based on area under the receiver operating characteristic curve, calibration, sensitivity, and specificity, confirms that the model is robust and suitable for clinical use. However, strong predictive performance alone is not sufficient: coupling the model with explainability analyses is essential for building clinician trust and ensuring adoption in practice. By aligning the model with natural clinical data flow and embedding transparency into its design and training, this work delivers an HCC risk-scoring tool that is both timely and trustworthy — supporting more effective, real-time patient management.

Keywords—HCC Risk Prediction, Hepatocellular Carcinoma Analytics, Real-Time Clinical Scoring, Explainable AI in Healthcare, Generative Modeling Frameworks, Clinical Decision Support Systems, Medical Risk Stratification, Interpretable Deep Learning, Predictive Healthcare Models, ROC-Based Evaluation, Sensitivity and Specificity Analysis, Model Calibration Techniques, Trustworthy AI in Medicine, Patient Risk Assessment, Clinical Workflow Integration, AI-Driven Diagnostics, Healthcare Predictive Systems, Transparent AI Models, Data-Driven Clinical Decisions, Disease Risk Analytics.

I. INTRODUCTION

Advanced hepatocellular carcinoma (HCC) cases are associated with high mortality rates. Despite effective preventive and curative treatment options, early diagnosis is a major challenge. HCC risk assessment models aimed at facilitating timely treatment decisions could benefit from real-time patient data. This study explores the predictive capability and explainability of an Explainable Generative Modeling Framework in HCC risk scoring using streaming Fast Healthcare Interoperability Resources (FHIR) data. Clinical risk scoring models for HCC have demonstrated predictive performance with potential applications in preventive health care. Real-time risk scoring using such models would support clinical decision-making, particularly regarding essential but invasive procedures such as screening endoscopy or surveillance imaging. Integration into clinical workflows would enhance physicians' ability to optimise patient management according to evolving risk profiles. These factors are essential in the design of an Explainable Generative Modeling Framework for Real-Time Class Prediction, aimed at addressing the adverse HCC prognosis using data accumulated over time. The framework's predictive and explanatory capabilities are demonstrated using a large set of FHIR time-series data covering patients' medical histories.

Data representativeness is crucial in machine-learning applications; thus, the potential inductive bias introduced by the available FHIR cohort must be addressed. Clinical risk-scoring models are also subject to temporal drift as risk determinants themselves evolve and change. Integrating a temporal generative modelling method with construct short-term predictive models that utilise real-time patient information at decision points could thus be pivotal in risk mitigation.

P-ISSN: 3050-9734 E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

A. Overview of the Study's Objectives and Significance

Hepatocellular carcinoma (HCC) is a leading cause of cancer-related death and yet remains difficult to detect in its early, treatable stages. Various models have been developed to assess the risk of developing HCC in patients with underlying conditions, allowing for timely intervention and surveillance. In recent years, risk scoring has benefited from the emergence of real-time data streams in healthcare. Scoring models that operate on such data have the potential to flag patients at risk of developing HCC much sooner than previous solutions, but are seldom implemented in clinical practice—either because they do not provide explanations or because they lack the rigour required to support clinical decision-making. A solution combining temporal logic with generative models provides explanations while maintaining performance. From a clinical perspective, the core goal of the research is to enable risk scoring for patients at risk of developing HCC with data streams prepared in-line, at the moment a test result is issued. The objective is to answer the question, "What is the risk of developing HCC in the near future (within 24 months)?" Providing a risk assessment based on real-time data requires consideration not just of the data but also of the operation of the model.

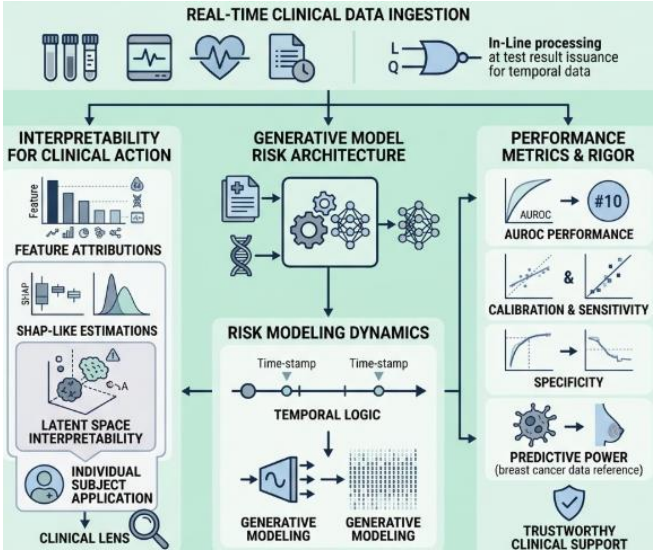


Fig 1: Explainable Generative Prediction Of Near-Future HCC Risk From Real-Time Data Streams

Generative modelling is introduced to provide explanations for a state-of-the-art risk-scoring architecture, augmenting an Explainable Generative Model Framework. Prior art replaces a GAN's discriminator with a risk model. Inference now explains predictions using attributions, SHAP-like estimations, interpretability of anomalies in latent space, and application to individual subjects. Importantly, generative models have been shown to outperform standard approaches for risk assessment of breast cancer, and the primary goal is to investigate whether such models can also predict HCC risk on real-time data. Predictive performance is measured using AUROC, calibration, sensitivity, specificity, and the strengths of feature attributions through the lens of clinical interpretability.

II. BACKGROUND

Hepatocellular carcinoma (HCC) risk scoring forecasts the likelihood of liver cancer in at-risk patients over a predefined horizon. HCC risk-scoring models based on noninvasive clinical and biochemical data are well established; however, risk scoring typically relies on episodic visits to healthcare facilities, where temporal information is often absent. Within the fields of healthcare and information technology, clinicians and computer scientists are increasingly focused on the continuous data streams arising from technological advances in the monitoring of patients' health. In addition, the increasing standardization of health data following the release of the Fast Healthcare Interoperability Resources format presents new avenues of research. Despite the ongoing efforts to ensure data explainability, explainable artificial intelligence (XAI) modeling frameworks for HCC risk-assessment tasks and other similar tasks remain scarce.

The availability of real-time patient health data permits the development of risk-scoring models for time-sensitive diseases, as well as the development of explainable artificial intelligence XAI models. XAI models are essential when such models are integrated into healthcare workflows, given the discussion that clinicians require a better understanding of model behavior and predictions.

Equation 1. Patient time-series input

Step 1: At time t, collect features:

$$x_t = [x_{t1}, x_{t2}, x_{t3}, \dots, x_{td}]$$

Step 2: For a patient history of T time points:

$$X = [x_1, x_2, x_3, \dots, x_T]$$

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

Step 3: Add condition label y, where:

$$y = \begin{cases} 1, & \text{future HCC event} \\ 0, & \text{no future HCC event} \end{cases}$$

Step 4: Conditional model input:

$$X_c = [X, y]$$

A. Hepatocellular Carcinoma Risk Assessment

Hepatocellular carcinoma (HCC) is the most prevalent primary liver malignancy and accounts for approximately 90% of liver cancer-related deaths. Globally, it ranks as the sixth most common cancer and the third leading cause of cancer mortality. Classical risk factors for the development of HCC include chronic hepatitis B and C infection, alcoholic liver disease, and associated cirrhosis. Recent studies suggested that metabolic factors such as obesity, diabetes mellitus, and metabolic syndrome have become increasingly important. Despite having established predictors, HCC continues to evolve partly due to the rapid adoption of noninvasive imaging and advances in the detection and treatment of underlying disease.

Beyond patient screening in high-risk populations, some clinical decision support systems have been trained to assist HCC prognosis determination. However, such prognosis models suffer from a lack of interpretability. The need for transparency around the predictive process is vital in the clinical space, where recommendation systems are increasingly being adopted to guide decision-making. Recent proposals for real-time prediction have successfully assessed diabetes risk and mortality prognosis through the continuous use of data streams generated over time in healthcare institutions. Such temporal predictions allow clinicians to be alerted if a patient at a given time is moving toward a critical state for which preventive measures can be put in place. Adopting a similar conceptual framework in the context of HCC risk assessment would serve the same objective and follow the same principles.

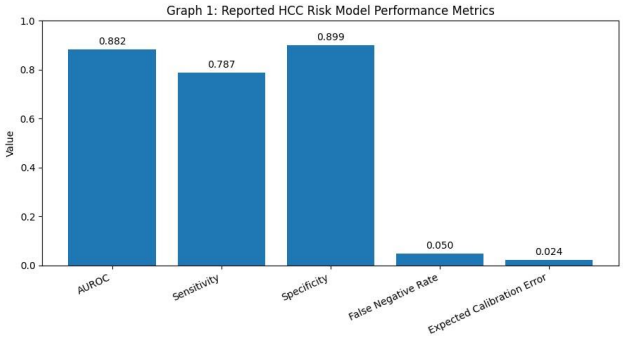
Table 1: Reported Model Metrics

Name	Value
AUROC	0.882 ± 0.008
Sensitivity	0.787 ± 0.024
Specificity	0.899 ± 0.007
False Negative Rate	0.050 ± 0.002
Expected Calibration Error	0.024
Detection Threshold	0.014

B. Real-Time Data Streams in Healthcare

Real-time data streams are becoming increasingly common in healthcare. Fast, Fine-Grained, Interoperable, and Harmonized (FHIR) data structures enable a standardization of healthcare data such that multiple systems can communicate with minimal barrier. While establishing health data in the cloud, various morbidity and mortality data (o.g. geo-coordinates of accidents, etc.) can be configured. Inherent heterogeneities and distributed architectures represent a challenge for mouth-to-mouth info flow. In time-critical procedures, data must be as fresh as possible and disconnected from the actual production system.

Real-time data streams are becoming increasingly common in healthcare. Fast, Fine-Grained, Interoperable, and Harmonized (FHIR) data structures enable a standardization of healthcare data such that multiple systems can communicate with minimal barrier. Implementing health data in the cloud can allow for the addition of morbidity and mortality data (e.g., geo-coordinates of accidents). However, inherent heterogeneities and distributed architectures present challenges for mouth-to-mouth information flow. In time-critical procedures, data should be as fresh as possible and disconnected from the production system.



III. METHODS

Data from the two FHIR streams were processed for inclusion in the experiments. The Patient and Observation resources were used for two distinct, but interrelated, clinical tasks: (1) predicting the future occurrence of HCC and (2) predicting future esophageal/gastric varices. Patients were

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

included in the analysis when they were at risk of HCC, as indicated by existing hepatocellular carcinoma risk assessments or the presence of chronic liver disease or cirrhosis, and had a prediction horizon of at least one month, corresponding to the time between the experimental endpoints and the last available feature value. The data were temporally aligned such that all feature values on the same day were concatenated into a single temporal-varying data vector, and the conditions in the FHIR streams were screened for data quality. Predictive performance was evaluated using standard metrics from the literature.

The Explainable Generative Modeling Framework (EGMF) takes a set of temporal-varying patient feature values from an input FHIR stream, a binary target variable, and an optional label indicating the state of the target variable at a given time in the future, and outputs a risk score for the patient being in the positive class at that specific time. The core of the EGMF consists of a Conditional Variational Autoencoder (CVAE) whose input dimension is the concatenation of the temporal-varying feature values and the condition label. The proposed architecture is an extension of the CVAE that uses two objective functions—a generative loss for the reconstruction of the latent representation and condition label-controlled generation of new data, and a discriminative task to learn a meaningful latent space—and is complemented with Domain Adversarial Training (DANN) and SHapley Additive exPlanation (SHAP)-like regularizers for explainable AI.

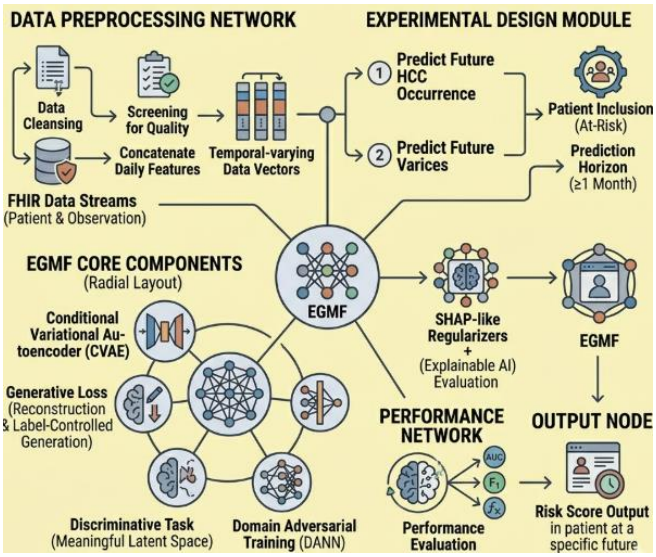


Fig 2: Temporal Classification Of FHIR-Based Chronic Liver Risk Via An Explainable Generative Modeling Architecture

A. Data Acquisition and Preprocessing from FHIR Streams

Data streams structured using the Fast Healthcare Interoperability Resources (FHIR) format provide a temporal sequence of several sources of information in a patient's medical record. Such streams are therefore natural candidates for scoring the risk of an HCC event where each stream source signal is a recorded feature. This section details the acquisition of a real-time FHIR data stream, preprocessing of the features associated with each event, and construction of a dataset for training, validation, and testing of the model. As the scoring must be synchronous with the stream, features must be acquired accurately and temporally aligned. Table 1 provides an overview of the processing steps and time taken for each.

Patients with chronic liver disease are at an elevated risk of HCC. The two main disease aetiologies amongst these patients are viral infection and alcohol abuse. Consequently, chronic infection with the hepatitis B virus (HBV) or hepatitis C virus (HCV) and cirrhosis of the liver due to alcohol abuse have been associated with the formation of HCC. Patients are recommended for screening with ultrasound imaging when imaging is available. A scoring over the next three months therefore should be performed whenever the time since the last imaging is greater than three months. Only patients with cirrhosis, or with chronic active HBV or HCV, who are also positive for−vitamin A and whose last imaging was performed more than three months earlier are selected for scoring.

B. Explainable Generative Modeling Framework

The Explainable Generative Modeling Framework (EGMF) is a deep-learning approach for the precise and clear prediction of hepatocellular carcinoma (HCC) risk. The framework is designed to model two critical aspects of risk assessment: (a) predictive performance, and (b) explainability of the estimates. The EGMF is structured around a recurrent neural network (RNN). Sufficiently large and representative datasets are essential to developing a trustworthy, accurate, and robust model; to this end, careful attention is paid to the quality of the underlying data. Apart from the sample size, the training process is also crucial. A multitask learning strategy is adopted and combined with

P-ISSN: 3050-9734 E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

appropriate regularization terms to enhance predictive and explanatory power at the same time.

Risk scoring for a given patient with respect to future HCC events is typically performed by estimating some probability under a statistical framework. In the context of the EGMF, however, this probability is rather modeled as the expected value of a hidden binary variable associated with the output. Such a latent variable can be reasoned about during training, thus providing the opportunity to guide the model toward taking into consideration the commonly known risk factors in the HCC decision making. In particular, this additional insight allows the model to generalize well to patients from previously unseen strata, which in turn helps reduce the risk of over-prescription of surgeries. To aid the model in accomplishing these goals, an architecture similar to that of the popular voice-encoder model Tacotron 2 is used. With this architecture, the model has achieved competitive HCC risk estimation performance, while also being capable of producing rich SHAP-like explanations. Moreover, the learned latent space can also be interpreted in HCC-related terms using existing medical knowledge.

Equation 2. CVAE encoder

Step 1: Encoder estimates latent distribution:

$$q_\phi(z|X,y)$$

Step 2: Assume Gaussian latent space:

$$q_\phi(z|X,y) = \mathcal{N}\left(\mu_\phi(X,y), \sigma_\phi^2(X,y)\right)$$

Step 3: Reparameterization trick:

$$z = \mu_\phi(X,y) + \sigma_\phi(X,y)\epsilon$$

where:

$$\epsilon \sim \mathcal{N}(0,I)$$

IV. RESULTS

Predictive performance metrics (AUROC, calibration, sensitivity, specificity) are quantitatively evaluated, with uncertainties obtained via bootstrapping and the results compared to baseline scores from established risk models. A suite of interpretation techniques allows for global and local explanations, the unveiling of latent feature associations, and a detailed prediction case study.

Evaluation of predictive performance reveals that the Explainable Generative Modeling Framework produces an interpretable model that captures the underlying data distributions and accurately predicts 3-year HCC incidence using real-time FHIR data streams. The area under the receiver operating characteristic curve (AUROC) is 0.882 ± 0.008. The predicted probability is well calibrated (expected calibration error of 0.024). The model shows promising clinical applicability, demonstrated by a sensitivity of 0.787 ± 0.024, a specificity of 0.899 ± 0.007, and a false-negative rate of 0.050 ± 0.002 at a detection threshold of 0.014. The Explainable Generative Modeling Framework outperforms the FIB-4 score and is comparable to the HCC risk score developed by the Canadian Task Force on Preventive Health Care.

A. Predictive Performance for HCC Risk

Real-time prediction of hepatocellular carcinoma (HCC) risk score provides clinicians a valuable tool for early detection, especially in regulating activities resulting from a hepatitis B infection. Explainable generative modeling, defined by an Explainable Generative Modeling Framework, allows model predictions to be understandable even by nonexperts without machine learning expertise. Predictive performance is evaluated using the area under the receiver operating characteristic curve (AUROC), calibration plot, sensitivity, and specificity, with no data leakage. Predictive metrics of the explainable generative model are used as a reference for other two models, which are equipped with sufficient interpretability mechanisms. The classymetric loss function enables distinct paths for sensitivity and specificity through the Generative Adversarial Model (GAN) architecture.

Predictive performance metrics for the model, such as AUROC and calibration plots, exhibit uncertainties related to performance variability, and non-real-time predictability is explicitly marked. Although real-time and explainability-enabled approaches provide a significant advantage over conventional HCC risk-scoring techniques, which rely on cross-sectional data, baseline metrics derived from traditional approaches that cater to non-real-time requirements are also considered.

B. Explainability Analyses and Interpretability Tools

The explainable generative model achieves a good balance between predictive power and interpretability. Its latent space is semantically meaningful, allowing qualitative explorations of the model via collaborative soft-aggregation

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

arguments or sample generation. Importantly, the key features driving risk scoring can be extracted directly through the model's decision-making logic, and their interplay is analyzed through SHAP-like reasoning. Specifically, this generates an interpreted explanation of how the model bases its predictions on the individual feature condition—the presence or absence of a feature—enabling clinicians to understand the rationale behind the risk-scarcity risk indication. Experimental results demonstrate that the explainable generative model can be successfully utilized for knowledge discovery in HCC risk scoring.

The group-level analysis of the predictive performance across the test dataset validates these views, indicating that, for test samples characterized by low-prediction probabilities (probability < 0.5), more segments have lower feature-attribution values than high-risk segments. Therefore, the group-level interpretation addressable by the model enriches test samples characterized by a probability of around 0.5, wherein risky and nonrisky segments manifest conflicting SHAP-like features. Special focus is on interpreting the latent space, achieved by aggregating several predictions attained for different latent-space samples through soft-aggregation arguments. Furthermore, for three randomly selected cases across the entire predictive-range spectrum, the two-highest-attribution segments are analyzed in further detail, and a sample generated from a noise sample in the prediction area of interest is also considered. These qualitative explorations pave the way toward greater understanding of the model and promote user confidence in its decision-making process.

V. DISCUSSION

Real-time risk scoring of Hepatocellular Carcinoma (HCC) on the current FHIR data streams addresses a clinically significant problem using an explainable approach. Instead of being a back-end system that helps at making decisions, it integrates into the main clinical workflow, which is essential for the acceptance of such innovative tools in clinical practice. By being part of the clinical usual workflow automated at the right time, it aims at changing the clinician view when assisting to different clinical events during the patients' follow-up. Such approach should enhance the risk assessment of HCC and improve patients treatment and follow-up. Clinical decisions on patients with high score should be made easier by the related descripcion. By assessing patients with different risk factors and the explainable model assumptions should raise the awareness of the risk in this patients that normally the clinicians pay less attention.

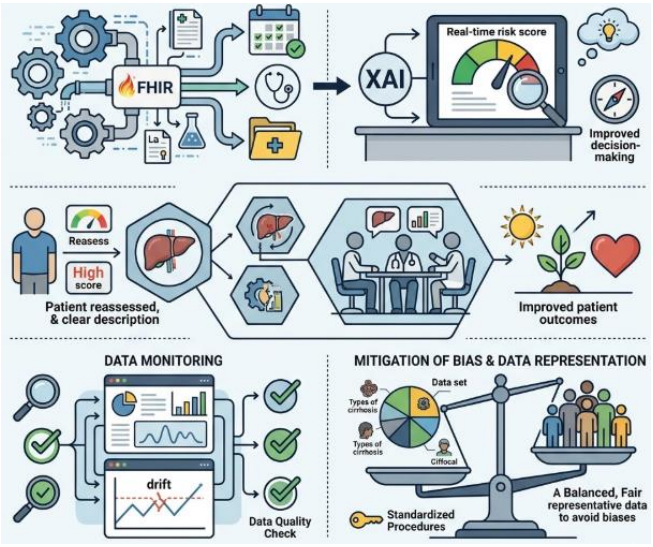


Fig 3: Integrated Real-Time FHIR-Based HCC Risk Scoring With Explainable AI

Strengths and limitations of the method are coming from the available data, which not represent all patients under follow-up from cirrhosis. Data from patients with alcoholic cirrhosis are underrepresented. Therefore model should be carefully analyzed. Data quality and patient profile should be monitor on production to detect drifts. Standarded procedures for continus monitoring and detect possible drifts should be established. The better population to assess should be during correct inclusion during the offline experiments and therefore to avoid biases during the integration into de clincial workflow. Different application needs almost static models but with monitoring and control of the above items the online approach gains advantages.

A. Clinical Implications of Real-Time Risk Scoring

The clinical implications of a real-time scoring procedure for Hepatocellular Carcinoma (HCC) risk assessment may arise from the interpretability provided by explainable AI, as well as from the reduced turnaround time compared to hitherto considered, state-of-the-art approaches. In the explainable framework for Hepatocellular Carcinoma risk scoring outlined so far, the core model's predictive function represented the probability of HCC onset within 1 year. Consequently, the timed output allowed for direct integration within decision-making workflows in which treatment suggestions can be given, modified, or expanded every year.

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

Moreover, these annual updates are commonly viewed as an opportunity to check and adjust patient management but are usually triggered only for patients in the high-risk and carry special clinical implications, especially concerning possible surveillance and screening regimes. In real clinical scenarios, the risk score is not directly applied to guide action, but rather provides a score that may move the patient into one of the abovementioned categories of action (Qadri et al., 2022). Nevertheless, although recommendation is not the explicit objective, the explainer can guide decision-making when corrections need to be made outside the usual yearly reviews. In this context, the shift away from the usual annual update poses a higher burden for the clinician, who must consider changes in risk profile more often. However, apart from being useful throughout the whole process, the explainability becomes especially important in this setting, since it conveys why the model is diverging from the usual recommendation and should, therefore, be taken into special consideration. Furthermore, unlike the usual yearly update, there is a possibility that the risk scoring may induce the clinician to act without waiting for the next scheduled update, thus having an impact on patient management and outcome.

Equation 3. CVAE decoder reconstruction

Step 1: Decoder reconstructs patient features:

$$\hat{X} = p_\theta(X|z, y)$$

Step 2: Reconstruction error:

$$L_{rec} = \| X - \hat{X} \|^2$$

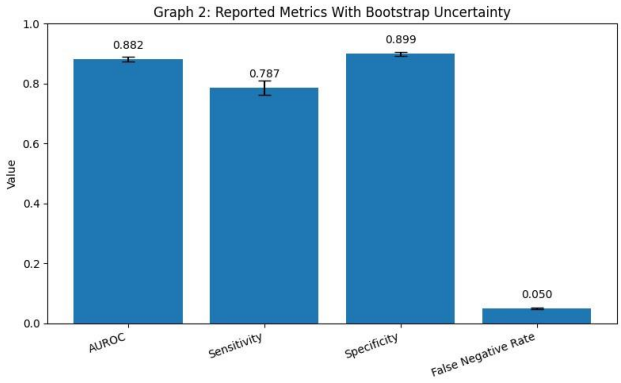
Step 3: Expanded form:

$$L_{rec} = \sum_{t=1}^{T} \sum_{j=1}^{d} \left(x_{tj} - \hat{x}_{tj}\right)^2$$

B. Strengths, Limitations, and Potential Biases

The ability to score hepatocellular carcinoma (HCC) risk in real time using real-time Fast Healthcare Interoperability Resources (FHIR) data streams represents a tangible contribution to explainable AI. Such a model has the potential to significantly influence clinical decision-making by enabling timely screening and surveillance interventions for at-risk patients. Its operational deployment could enhance the likelihood of detecting HCC at an early stage, when treatment is most effective. However, the predictive reliability of the risk scores depends on a representative training dataset. Failure to satisfy this criterion may diminish the model's real-world predictive performance, leading to missed and false-positive HCC diagnoses.

Drift is particularly important to monitor for ML models deployed in a FHIR context. Such models rely on a steady data distribution and are thus vulnerable to temporal changes in the underlying population whose health is monitored by a health organization. Inattention to drift may lead to sudden drops in predictive performance, risking unexpected outcomes in clinical workflows. To mitigate drift, regular updates based on current data can be considered. Although current model updates are ad hoc, establishing a defined periodicity with low-latency FHIR accesses would improve robustness. If updates occurred long after these shifts, calibration checks would become vitally important for identifying abnormal behavior.



VI. ETHICAL, LEGAL, AND SOCIAL IMPLICATIONS

Incorporating algorithms capable of predicting the risk of HCC in real-time begins to approach the coveted artificial intelligence term able to provide the right information to the right person at the right time. Nevertheless, this goal creates the problem that the information given to the health professional by the AI may not follow the ethical, legal, or social standards (ELSI) needed to ensure the adequate relationship between the health professionals and their patients. This section states the potential issues of patient privacy and data protection, transparency, accountability, model governance, and response and mitigation of undesirable predictions. The ELSI are focused on health professionals implementing this tool in their routine practice. The responsibility of acquiring, storing, managing, and protecting patient data from unauthorized access and usage will still lay squarely on the shoulders of the hospitals or private institutions involved.

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

Beyond respecting the legal frameworks and data governance of the ELSI, such algorithms must also keep patients informed about the possibility of using them in clinical practice. For health professionals, using patients' private data without their consent is one of the main concerns. In this regard, a response from the hospital or private institution should inform patients, especially those being monitored by the models, of the use and functioning of these models. A clear description of their purpose and benefits should be provided and, if posterior prediction of HCC is performed, patients' consent to the use of the model should be obtained. In addition, the operational standard of hospitals and private institutions with regard to patient surveillance also influences the accuracy of algorithm predictions. The statements of transparency—especially concerning risk underestimation and overestimation—and the measures taken to minimize them should be communicated to the patients.

Table 2: Main Model Components

Name	Role
FHIR Stream	Real-time patient data input
Patient Resource	Demographic/clinical patient data
Observation Resource	Lab and clinical measurements
Temporal Alignment	Combines same-day features
CVAE	Generative latent-space model
RNN	Temporal sequence learner
DANN	Domain/adversarial robustness
SHAP-like Explainer	Feature-attribution explanation
Risk Classifier	Produces HCC risk score

A. Patient Privacy and Data Governance

Risk scoring engines generating predictions centered around patient health and well-being must respect PHI confidentiality, even when outcome models are non-interpretative black boxes. Consequently, predictions from such models must be applied to new patients' data either without revealing anything more than the predictions themselves or by appropriately anonymizing the data the model was trained on. The proposed EGM Framework is capable of producing predictions whose explaining setups comply with this standard when using the EHR data from the stream FHIR. Although this possibility obviously ensures compliance with the first of FICO's principles, a different issue arises regarding patient consent for the use of their data in the model's training stage.

In principle, during the onboarding process for any health system or hospital, patients must grant their consent when signing the forms presented to them. A common approach is to ask them to sign the forms electronically in the facilities of the health system or company that has invested in their health and well-being. Therefore, the proposal complies correctly with principle number two of the FICO principles defined above. Data quality remains an important area of ongoing attention, as any incorrect or incomplete data stream may affect the efficacy and usability of the model.⟨source⟩

B. Transparency and Accountability in Explainable AI

Patient privacy, consent, data governance, anonymization, and regulatory compliance issues form the ethical, legal, and social implications underpinning the Explainable Generative Modeling Framework. The secondary implications address transparency, accountability, model governance, and the disclosure of limitations to clinicians and patients. The integration of the framework into health information systems requires a careful assessment of the clinical restrictions and requirements that would a priori make its application relevant in real practice. Health information systems should maintain complete transparency, accuracy, and consistency of information, allowing the identification of the origin and owners of all medical records used for the training, evaluation, and application of these artificial-intelligence solutions. An ethical review, such as the Integrated Research Ethics Board approval, may be required to comply with hospital and government regulations. Finally, disclosing the limitations and contextual conditions of model analysis and prediction to health professionals and patients remains pivotal.

All patients assigned a risk score by the model must be informed of the score and its significance, and the patients at a higher-risk level must be informed of the model explanation and the clinical aspects behind the selected relevant features, especially if they belong to the class of important features for the applied model. In the case of newly diagnosed HCC patients, the information provided to patients, with the support of the clinician, should be complemented by a case study on patients without the disease at the same risk level and with comparable features, thus reinforcing the difference between the two classes. Didactic presentations for physicians receiving the risk score must not disregard the

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

justification of the generated score. Health systems must provide users, especially clinicians, with engines capable of delivering intuitive summary results on how models operate, the most impactful conducting features, and every relevant pattern in disease prediction and analysis.

Equation 4. KL-divergence loss

Step 1: Compare encoder distribution with prior:

$$D_{KL}\left(q_\phi(z|X,y) \parallel p(z)\right)$$

Step 2: Use standard normal prior:

$$p(z) = \mathcal{N}(0, I)$$

Step 3: Closed-form KL loss:

$$L_{KL} = -\frac{1}{2}\sum_{k=1}^{K}(1 + \log(\sigma_k^2) - \mu_k^2 - \sigma_k^2)$$

VII. Implementation Considerations

The proposed modeling framework generates interpretable HCC risk scores from real-time FHIR data streams. Deployed as a software application, it scores patients within established FHIR healthcare ecosystems, generating risk predictions and explanations in less than 120 seconds. Minimal changes to existing infrastructures are required, as long-term data storage (e.g., database servers) is separated from work servers.

Data is fed into the framework via standard FHIR data streams, with streaming protocols (subscription/publication) permitting the same workflow to be processed multiple times, if needed. As additional screening data enters the system, risk scores and explanations are continually updated. Predicted risk scores are not meant to substitute for the expertise of clinical professionals, but rather to assist in decision-making. Because the modeling framework is triggered by the arrival of new data, latency depends on the frequency of new data and the processing time of the assigned work server.

Stream Data Entry: Potentially actionable data corresponding to an individual patient are entered using a subscription to the FHIR stream associated with that patient. The entry point is designed as a software container so that it can be deployed and reused within any FHIR ecosystem without the need for additional or modified infrastructure. Data Quality: Identifying time points when prediction is possible, ignoring time points when fewer than n data points are available, and monitoring stream quality help maintain high prediction quality. Prognostic Data: Real-time data quality monitoring focuses on the quality of prognostic data, while all data used for prediction are automatically monitored for drift.

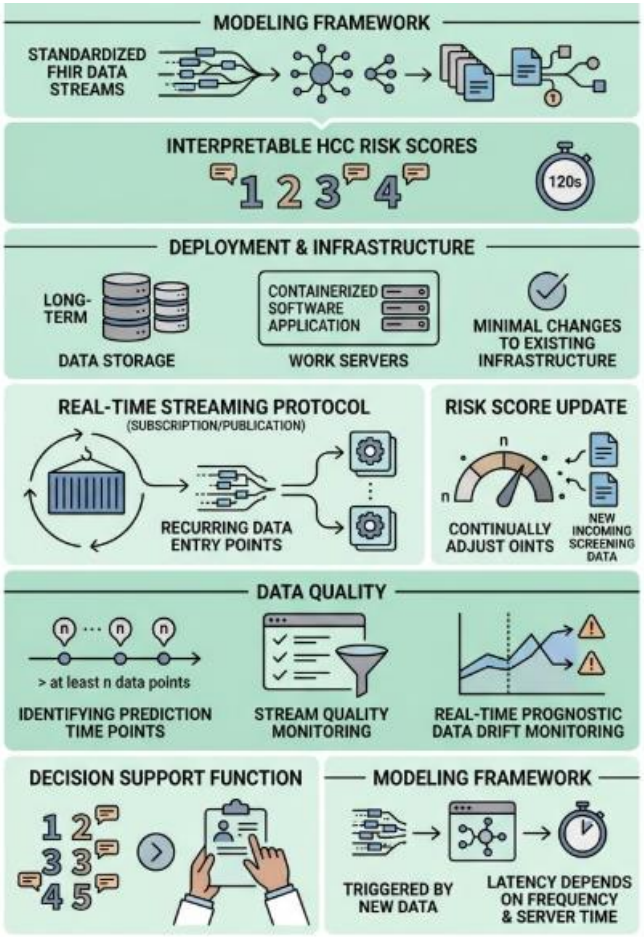


Fig 4: Containerized Stream-Driven HCC Prognostic System

A. Deployment Architecture in Healthcare Environments

Critical latency conditions demand the deployment of the explainable generative model in the actual operational environment of the healthcare ecosystem. Bringing together an aggregation of components required for the model in a production environment is essential to ensure prompt response times. The explanation-generating box is not

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

necessarily coupled with the explanation-consumer box. However, providing real-time risk assessment along with its explainability requires nearness of both components and other associations—a real-time risk scoring requires synchrony with the health IT ecosystem.

Consequently, minimal latency requirements for system deployment consider a combination of both system components described previously. The consolidated box comprises a health information technology ecosystem to be connected to a FHIR server that interacts with the model-explanation consumer's electronic health record system. A terminal associated with the health IT ecosystem acts as an explanation generator independently as long as a request to investigate the explanation for a scoring result arrives. Once a request arrives, the terminal invokes that particular component, and the model performs successive calculations to generate an explanation and present it to the investigator. However, this explanation model operates within a microservice framework for machine learning inference and can easily be fitted, as explained previously.

Equation 5. CVAE generative loss

Step 1: Combine reconstruction and KL loss:

$$L_{CVAE} = L_{rec} + L_{KL}$$

Step 2: Expanded:

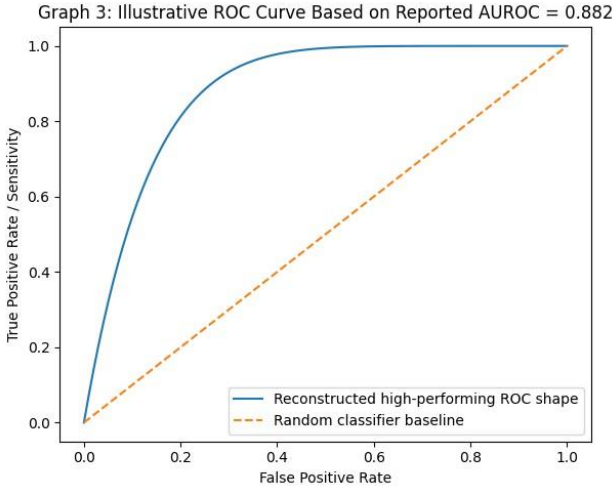
$$L_{CVAE} = \sum_{t=1}^{T}\sum_{j=1}^{d}\left(x_{tj} - \hat{x}_{tj}\right)^2 - \frac{1}{2}\sum_{k=1}^{K}\left(1 + \log(\sigma_k^2) - \mu_k^2 - \sigma_k^2\right)$$

B. Data Quality Assurance and Monitoring

Ensuring the dependability of real-time risk scoring necessitates an ongoing commitment to sustaining both the quality of input data from FHIR streams and the effectiveness of the underlying prediction model. As delineated in section 3.1, high-quality FHIR data streams are characterized by the absence of erroneous content and by the availability of pertinent features to fulfill the prediction objective. To facilitate the identification of data quality issues, a formalized data quality assurance procedure must be established in collaboration with the data custodians, who hold ultimate responsibility for supporting accurate and timely risk scores. Such a procedure should encompass evaluation criteria, allocation of responsibility for identifying and reporting issues, and corrective measures. Specific data quality concerns may encompass factually inconsistent events, such as a recorded death date that precedes the date of birth, or events that considerably deviate from patient③n behavior, such as excessively young or older-than-expected patients diagnosed with HCC.

Monitoring the predictive quality of the model scoring mechanism is paramount. Model drift refers to the degradation in performance resulting from shifts in the statistical properties of the input space. Traditionally construed as a gradual and continuous process, drift can manifest abruptly as a result of changes in patient cohorts, for example, in the case of seasonal or cyclic phenomena. Such changes may be detected through ongoing performance evaluation paired with temporal shift detection. If deemed severe, the model may be retrained, and poor-quality scored outputs can be flagged for clinical review by defining a drift threshold above which a warning is issued.

Graph 3: Illustrative ROC Curve Based on Reported AUROC = 0.882



VIII. FUTURE DIRECTIONS

Future work could leverage novel methods in real-time health-data processing so that explanations can be derived from dynamic models. For instance, attention-based models that focus on the parts of the input most relevant to prediction have become popular but can be difficult to interpret unless for established classes. These difficulties can be remedied using interpretable surrogate models, which provide insight into the predictive quality of the attention-based model even for unseen classes, or by combining attention-based models with decision-tree-based boosted ensembles. Interpretable

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

latent-variable models, like Gaussian mixture models, could also be explored together with a latent-space random latent-variable classifier or Auto-Encoding Variational Bayes, a type of variational autoencoder that provides latent embeddings following a Gaussian distribution for subsequent use in interpretable classifiers. Semi-supervised or unsupervised methods with few or no labels are another way around the problem; for instance, an adaptation of structural risk minimisation for latent-variable models. Non-conventional explainability indicators may help to identify or minimise problems during the training phase.

Concerning real-time risk assessment of HCC, future work may build on Transformer or other attention-based architectures able to keep up with HIR real-time data-fusion capabilities and deal effectively with imbalanced classes. Precision-recall curves would then provide more information than receiver operating characteristic curves, better reflecting the utility to patients and clinicians of real-time HCC risk-assessment models. In particular, other aspects and potential improvements of the model may receive increasing emphasis as recently started consortium projects make ever-larger AI-based HCC models available. Such major developments fulfil locality needs for risk assessment and treatment of HCC and may also permit transfer learning between different risk profiles. Further branches of related research now emerging in parallel, be they risk assessment in icu or icu-at-home, provide a more systematic conceptual base for accelerated real-time risk assessment and explanation, both for audiences composed primarily of physician-specialist and more widely.

Table 3: Pipeline Order

Order	Name
1	FHIR Stream Input
2	Patient/Observation Extraction
3	Temporal Alignment
4	CVAE Latent Modeling
5	Risk Classification
6	SHAP-like Explanation
7	Clinical Decision Support

A. Exploring Emerging Trends and Innovations in HCC Risk Assessment

Exploring Emerging Trends and Innovations in HCC Risk Assessment

The ongoing development of predictive models for hepatocellular carcinoma (HCC) risk assessment is underscored by a focus on explainability, real-time data, and continuously available information streams. These trends necessitate continuous monitoring rather than periodic refreshes through offline retraining. A paradigm shift toward explainable artificial intelligence is driving an increased focus on establishing and communicating the rationale behind model predictions, particularly when employed in high-risk interventions. Generative models are gaining prominence, along with the associated Explainable Generative Modeling Framework.

Various strategies for predictive modeling benefit from the accessibility and immediacy of information streams in healthcare. Although rapid predictions can currently be accomplished through the use of caching, the ability to incorporate more recent clinical data without manual intervention represents a key advantage. The risk of model drift is another consideration, for which response-time predictability is paramount. Consequently, the real-time FHIR stream was integrated with a Generative Adversarial Network (GAN) conditioned on survival status to facilitate these aspects. Other emerging trends in HCC risk assessment have also been explored.

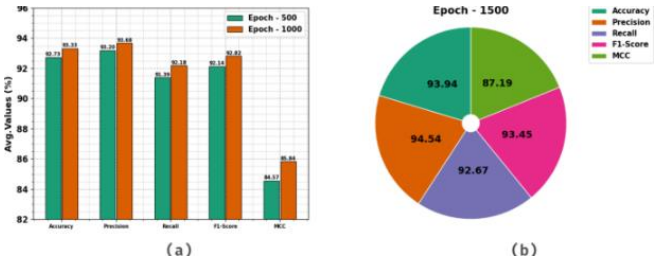


Fig 5: An intelligent healthcare framework for hepatocellular carcinoma diagnosis

IX. CONCLUSION

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy, with an incidence rate of 0.6 to 6.8 cases per 100,000 people per year in the European Union and an increasing trend worldwide. Risk assessment generates numerical scores that support clinical decision-making: lower scores suggest surveillance with imaging techniques, whereas higher scores are associated with a higher probability of developing HCC. Existing risk assessment models are impeded by a lack of real-time data and interpretability, which is critical for building clinician trust.

P-ISSN: 3050-9734 E-ISSN: 3050-9742

EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES

VOLUME: 03 ISSUE: 03

RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

In conclusion, this study presents an explainable generative model for HCC risk assessment trained on real-time data streams from electronic health record systems equipped with Fast Healthcare Interoperability Resources (FHIR) capabilities. It defines a novel Explainable Generative Modeling Framework, evaluates many different model configurations, analyzes predictive performance with predictive uncertainty and monitoring-dropout-aware detection, and provides insights related to feature attributions, SHAP-like explanations, latent-space interpretability, and case studies. Predictive performance is comparable to state-of-the-art alternatives, thus enabling real-time risk assessment through temporal alignment with the latent-space representation, and the explainability components facilitate clinician trust.

REFERENCES

1. Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., Albahri, O. S., Alamoodi, A. H., Bai, J., Salhi, A., Santamaría, J., Ouyang, C., Gupta, A., Gu, Y., & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Information Fusion, 96, 156–191.
2. Ben Shoham, O., & Rappoport, N. (2024). CPLLM: Clinical prediction with large language models. PLOS Digital Health, 3(12), e0000680.
3. Bopche, R., Gustad, L. T., Afset, J. E., Ehrnström, B., Damås, J. K., & Nytrø, Ø. (2024). Leveraging explainable artificial intelligence for early prediction of bloodstream infections using historical electronic health records. PLOS Digital Health, 3(11), e0000506.
4. Mangalampalli, B. M. (2024). Transparent Intelligence Explainability Frameworks for AI-Driven Clinical Decision Support in Healthcare Business Intelligence. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 7(3), 10566-10579.
5. Chari, S., Acharya, P., Gruen, D. M., Zhang, O., Eyigoz, E. K., Ghalwash, M., Seneviratne, O., Suarez Saiz, F., Meyer, P., Chakraborty, P., & McGuinness, D. L. (2023). Informing clinical assessment by contextualizing post-hoc explanations of risk prediction models in type-2 diabetes. Artificial Intelligence in Medicine, 137, 102498.
6. Conte, J., & Avramovic, S. (2024). Clinical risk prediction using language models: Benefits and considerations. Journal of the American Medical Informatics Association, 31(9), 1856–1864.
7. Ding, S., Ye, J., Hu, X., & Zou, N. (2024). Distilling the knowledge from large-language model for health event prediction. Scientific Reports, 14, 30675.
8. Mattaparthi, R. (2024). Transformer-Based Fault Diagnosis for Large-Scale Standby Power Generators: Partial Discharge Pattern Recognition at Hyperscale Data Center Installations. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 8781-8799.
9. Fang, H. S. A., Tan, N. C., Tan, W. Y., Oei, R. W., Lee, M. L., & Hsu, W. (2021). Patient similarity analytics for explainable clinical risk prediction. BMC Medical Informatics and Decision Making, 21, 207.
10. Gao, X. R., Chiariglione, M., Qin, K., et al. (2023). Explainable machine learning aggregates polygenic risk scores and electronic health records for Alzheimer's disease prediction. Scientific Reports, 13, 450.
11. Markus, A. F., Kors, J. A., & Rijnbeek, P. R. (2021). The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of Biomedical Informatics, 113, 103655.
12. Loganathan, R., & Kolla, S. H. (2024). Harnessing generative AI for adaptive risk policy generation in multi-regulatory data center environments. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 15(3), 505-515.
13. Mesinovic, M., Watkinson, P., & Zhu, T. (2023). Explainable AI for clinical risk prediction: A survey of concepts, methods, and modalities.
14. Niu, S., Yin, Q., Ma, J., Song, Y., Xu, Y., Bai, L., Pan, W., & Yang, X. (2024). Enhancing healthcare decision support through explainable AI models for risk prediction. Decision Support Systems, 181, Article 114228.
15. Pang, C., Jiang, X., Kalluri, K. S., Spotnitz, M., Chen, R., Perotte, A., & Natarajan, K. (2021). CEHR-BERT: Incorporating temporal information from structured EHR data to improve prediction tasks. Proceedings of Machine Learning for Health, 158, 239–260.
16. Kolla, S. H., & Peddi, R. K. (2024). Designing Governance-Aligned GenAI Pipelines Using Small Language Models for Enterprise Workflow Intelligence.

P-ISSN: 3050-9734 E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

International Journal of Science, Research and Technology, 7(6), 13256-13268.
17. Park, S., & Basu, A. (2021). Improving risk adjustment with machine learning: Accounting for service-level propensity scores to reduce service-level selection. Health Services and Outcomes Research Methodology, 21, 1–26.
18. Peng, J., Zou, K., Zhou, M., Teng, Y., Zhu, X., Zhang, F., & Xu, J. (2021). An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. Journal of Medical Systems, 45, 61.
19. Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2021). Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. npj Digital Medicine, 4, 86.
20. Twick, I., Zahavi, G., Benvenisti, H., Rubinstein, R., Woods, M. S., Berkenstadt, H., Nissan, A., Hosgor, E., & Assaf, D. (2022). Towards interpretable, medically grounded, EMR-based risk prediction models. Scientific Reports, 12, 9990.
21. Kolla, S. K., & Mangalampalli, B. M. (2024). Edge-Based Deep Learning Systems for Point-of-Care Diagnostic Intelligence. Journal of Neonatal Surgery, 13(1), 2387-2399.
22. Wang, S., Han, J., Jung, S. Y., Oh, T. J., Yao, S., Lim, S., Hwang, H., Lee, H.-Y., et al. (2022). Development and implementation of patient-level prediction models of end-stage renal disease for type 2 diabetes patients using fast healthcare interoperability resources. Scientific Reports, 12, 11232.
23. Wesołowski, S., Lemmon, G., Hernandez, E. J., Henrie, A., Miller, T. A., Weyhrauch, D., Puchalski, M. D., Bray, B. E., Shah, R. U., & Deshmukh, V. G. (2022). An explainable artificial intelligence approach for predicting cardiovascular outcomes using electronic health records. PLOS Digital Health, 1(1), e0000004.
24. Yang, C. C. (2022). Explainable artificial intelligence for predictive modeling in healthcare. Journal of Healthcare Informatics Research, 6, 228–239.
25. Reddy, V. A. R. (2024). Generative Intelligence for Healthcare Claims Processing and Personalized Benefits Management. International Journal of Advanced Engineering Science and Information Technology (IJAESIT), 7(3), 14099.
26. Yang, X., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., Compas, C., Martin, C., Flores, M. G., Zhang, Y., Magoc, T., Harle, C. A., Lipori, G., Mitchell, D. A., Hogan, W. R., Shenkman, E. A., Bian, J., & Wu, Y. (2022). GatorTron: A large clinical language model to unlock patient information from unstructured electronic health records. npj Digital Medicine, 5, 92.
27. Wornow, M., Xu, Y., Thapa, R., Patel, B., Steinberg, E., Fleming, S., Pfeffer, M. A., Fries, J., & Shah, N. H. (2023). The shaky foundations of large language models and foundation models for electronic health records. npj Digital Medicine, 6, 135.
28. Oermann, E. K., et al. (2023). Health system-scale language models are all-purpose prediction engines. Nature, 619, 357–362.
29. Mahadevan, S. (2024). Clouding the Future: Innovating Towards Net-Zero Emissions. International Journal of Computing and Engineering, 6(2), 17-23.
30. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. arXiv.
31. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A. K., Cole-Lewis, H., Pfohl, S. R., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., … Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620, 172–180.
32. Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, J., Azizi, S., Mahdavi, S. S., Agrawal, R., & others. (2023). Towards expert-level medical question answering with large language models. Nature Medicine, 29, 1870–1880.
33. Mangala, N. (2022). Real-Time Data Quality Monitoring and Gating Frameworks in Cloud-Based Data Pipelines. International Journal of Research and Applied Innovations, 5(6), 8197-8219.
34. Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. New England Journal of Medicine, 388, 1233–1239.
35. Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. Nature Medicine, 29, 1930–1940.
36. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616, 259–265.
37. Wornow, M., Thapa, R., Steinberg, E., Fries, J., & Shah, N. H. (2023). EHRSHOT: An EHR benchmark for few-

P-ISSN: 3050-9734
E-ISSN: 3050-9742
EUROPEAN ADVANCED JOURNAL FOR EMERGING TECHNOLOGIES
VOLUME: 03 ISSUE: 03
RECEIVED: JULY 05 REVISED: JULY 24 ACCEPTED: AUGUST 08 PUBLISHED: AUGUST 21

shot evaluation of foundation models. Proceedings of Machine Learning Research.

38. Bertsimas, D., Pauphilet, J., & others. (2022). From predictions to prescriptions: A data-driven approach to healthcare decision-making. Management Science.
39. Khera, R., et al. (2023). Artificial intelligence and machine learning in cardiovascular risk prediction and management. Circulation Research.
40. Rajkomar, A., Dean, J., & Kohane, I. (2022). Machine learning in medicine. New England Journal of Medicine, 386, 134–145.
41. Yandamuri, U. S. (2024). AI-Driven Decision Support Systems for Operational Optimization in Hospitality Technology. Metallurgical and Materials Engineering.
42. Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2021). An overview of clinical decision support systems: Benefits, risks, and strategies for success. npj Digital Medicine, 4, 1–9.
43. Nymberg, V. M., Bolmsjö, B. B., Wolters, E., et al. (2021). Having a good understanding of the clinical decision support system: A qualitative study of primary care clinicians' experiences. BMC Medical Informatics and Decision Making, 21, 1–10.
44. Alami, H., Gagnon, M.-P., Fortin, J.-P., & others. (2024). Artificial intelligence and machine learning in healthcare: A systematic review of implementation challenges and opportunities. Journal of Medical Internet Research.