ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 2 Issue 2, June 2022 / Page No: 183-191
Paper Id: JETA-V2I2P121 / Doi: 10.56472/25832646/JETA-V2I2P121

Original Article

Foundation Models for Multi-Modal Clinical Decision Support Systems

Vibhor Pal
Independent Researcher.

Abstract: *Decision support systems (DSS) are computer programs founded on artificial intelligence (AI) techniques that help in arriving at the right conclusion within a usually closed area of concern. One such DSS is called a clinical decision support system (CDSS) that can be utilized by clinicians in clinics and hospitals. This paper explores the effectiveness of different machine learning (ML) and deep learning (DL) models on Multi-Mode Clinical Decision Support Systems (CDSS) on the MIMIC-IV data set. Conventional approaches, such as the Random Forest (RF), demonstrated an accuracy of 77% whereas DL algorithms, such as LSTM or TieNet, demonstrated 82.01 and 84.8% rates, respectively. A hybrid XGBoost + Decision Tree model was suggested to improve the predictive performance, and the maximum accuracy was 92%. High precision (96%), recall (95%), and F1-score (95%) also showed that the model is robust and reliable to help make clinical decisions. These findings highlight the usefulness of hybrid ensemble methods that can be used to take advantage of multi-modal clinical data to make accurate and interpretable predictions in CDSS.*

Keywords: *Automated Machine Learning (Automl), Deep Learning, Intensive Care Unit, Mortality, Multimodal, Feature Fusion.*

I. INTRODUCTION

Clinical care is a critical factor in the process of patient health outcomes, which includes diagnosis, treatment and follow-up. The successful performance of these processes demands the involvement of medical practitioners with broad knowledge and expertise. Nevertheless, the healthcare field of the world is experiencing increased pressures, such as demographic changes, a rise in the rate of chronic illnesses, and continuous resource limitations. In the United States, the annual healthcare spending is more than $1.7 trillion US dollars [1]. These issues highlight the problems of the necessity of sophisticated clinical decision-making tools that guarantee the efficient, accurate, and high-quality care of patients [2]. Clinical Decision Support Systems (CDSSs) are smart software programs that have been developed to aid health care providers through real-time, evidence-based suggestions based on complicated medical data. The traditional CDSSs have, to a large extent, referred to limited data through a single mode, e.g., textual records or medical images processed on their own. This small focus restricts their capability to define the complexity of patient health data and impairs their performance in the contemporary data-intensive clinical environment [3][4].

In order to overcome these shortcomings, multimodal clinical diagnosis combines many different types of datasets, such as EHRs, MRIs, and patient documents, to construct a complete vision of patient conditions [5][6]. Here, the reporting and classification of diseases are some of the tasks that depend on both visual and textual data to generate the correct diagnoses. Nonetheless, the generation of medical reports of the imaging modalities, such as the chest X-rays (CXR) and the computed tomography (CT) scan, is still a tedious task that is repetitive and can contribute to clinician burnout and associated human error. Workload can be greatly reduced through the implementation of automated report generation systems, which result in preliminary drafts that are reviewed by clinicians, thereby reducing workflows and minimizing the error rate, thereby improving the general clinical performance [7][8].

As Artificial Intelligence (AI) is rapidly developing, foundation models have become a revolutionary way to approach healthcare applications [9]. In contrast to AI systems based on traditional machine learning (ML), foundation models are trained via large-scale pre-training on diverse data to build generic models that can be tailored to other medical procedures in the future. This type of model is especially appropriate to the healthcare sector, where clinical decision-making is multimodal by nature and involves combining structured (including lab results and vitals) and unstructured (including clinical notes and medical images) data [10][11]. Based on these developments, the Multimodal Decision Support System (MDSS) is a new framework that integrates both structured and unstructured medical information and facilitates clinical reasoning and medical accuracy. With the potential to integrate the data of the patient health, MDSS provides a more holistic perspective of patient health and creates a better and more informed and data-driven decision-making system and can help improve the upcoming generation of intelligent, multimodal Clinical Decision Support Systems (CDSSs) [12].

A. Motivation and Contribution

The rationale of this research is that increasing complexity and volume of clinical data in healthcare make it harder to achieve accurate and timely decision-making. Conventional frameworks tend to be inadequate when it comes to the comprehension of the complex associations between multi-modal data, such as demographic data, vital signs, lab tests, and medical history. It is urgently necessary to have more robust, interpretable, and high-performing models that may assist clinicians in forecasting important outcomes, including ICU readmissions, with more accuracy. This study overcome these shortcomings by employing a hybrid XGBoost + Decision Tree model, which more predictive but maintain the interpretability of the model, thereby resulting in improved patient care and clinical efficiency. This study provides some of the important contributions as enumerated below:

- Employed DenseNet121 pretrained on ChestX-ray images to extract 1024-dimensional feature embeddings for integrating imaging data into the CDSS framework.
- Developed a comprehensive pre-processing pipeline involving missing value handling, text normalization (lowercasing), and feature concatenation for model readiness.
- Proposed a hybrid XGBoost + Decision Tree model combining accuracy and interpretability for clinical prediction tasks.
- Evaluated the model's overall performance as measured by measures like ROC-AUC, F1-score, recall, precision, and accuracy.
- Proved that hybrid ensemble models are effective in a multi-modal clinical data environment, which enables effective and sound decision support in healthcare systems.

The rationale behind the suggested Hybrid XGBoost + Decision Tree model is that the model address the shortcomings of one-model methods by integrating interpretability and excellent predictions. It is novel because of the combination of boosting and tree-based learning in one framework that allows handling heterogeneous clinical data of multiple modalities effectively. This design not only enhances model transparency and generalization but also introduces a more explainable and robust method for clinical decision support, contributing to safer and more data-driven healthcare outcomes.

B. Organization of the Paper

The remainder of this paper is organised as follows in Section II present background information and previous research. The study methodology is described in Section III. Section IV presents the propose results and comparison, conclusion with findings summarised in Section V.

II. LITERATURE REVIEW

A thorough review and analysis of significant research studies on Clinical Decision Support Systems was undertaken to guide and strengthen the development of the present study.

Siciarz et al. (2021) developed and studied an innovative ML-based clinical decision-support system. The model that performed best, as evaluated using double-nested cross-validation and the area under the curve (AUC), was selected for further research. To make sense of the model's forecasts, we used interaction values from Shapley's additive explanation (SHAP). The AUC was 0.98 ± 0.02, and the testing data accuracy was 93.8 ± 4.1% for Logistic Regression, making it the most effective method. The ΔD99% metric for PTV had the greatest impact on the model predictions, as revealed by the SHAP investigation. The left and right cochleae's ΔDMAX was the least significant characteristic [13].

Qjidaa et al. (2020). The input Images to the proposed model are pre-processed before further data is added. In order to do classification and deep learning (DL), the VGG CNN was utilized. In internal validation, the suggested model was 92.5% accurate, while in external validation, it was 87.5% accurate. Both the internal and external validations yielded values of 97% and 95%, respectively, for the AUC criteria. Internal validation yielded a sensitivity criterion score of 92% and external validation yielded a value of 87% [14].

Hamdaoui et al. (2020) offered a clinical support system to aid in the prediction of cardiac problems, which would aid in the diagnosis process and allow for improved decision-making by physicians. This work utilizes ML techniques such as Decision Tree, Random Forest, K-Nearest Neighbours, Support Vector Machine, and Naïve Bayes to forecast the onset of heart disease by mining data on risk markers from patient records. With a cross-validation accuracy of 82.17% and a train-test split accuracy of 84.28%, Naïve Bayes outperforms in predicting HD using the UCI data set, according to many studies [15].

Yahyaoui et al. (2019) evaluated DL methods in comparison to traditional ML methods. Two of the most well-known classifiers, Random Forest (RF) and Support Vector Machine (SVM), are instances of traditional ML techniques that they examined. The proposed method was assessed with the use of the open-source Pima Indians Diabetes database, which has 768 samples with 8 qualities each. Two hundred sixty-eight individuals with diabetes and 500 healthy controls were included in the sampling. Overall, RF obtained 83.67% accuracy, SVM 65.38 %, and DL 76.81 %. Results from the experiments demonstrate that RF outperformed DL and SVM approaches for diabetes prediction [16].

Andrianto et al. (2019) prioritize the identification of typhoid fever disease (TFD) due to the high prevalence of antibiotic resistance and the difficulty in diagnosing this endemic disease in Indonesia. The three supervised classification techniques utilized by this system are NB, kNN, and SVM. After that, the system was assessed and compared. When compared to the other two approaches, kNN's 88.7% accuracy was the best for typhoid fever illness classifications [17].

The review of related studies (as shown in Table I) reveals that most existing works use traditional or DL models on limited datasets with repetitive pre-processing and feature methods, leading to poor generalization. Few studies focus on hybrid or explainable models. Thus, there remains a research gap in developing a generalized, interpretable, and high-performing model suitable for diverse medical data.

Table 1 : Recent Studies on Clinical Decision Support Systems Using Machine Learning Techniques

Author	Proposed Work	Dataset	Results	Limitations & Future Work
Siciarz et al. (2021)	Developed a clinical decision-support system using multiple ML models evaluated with double-nested cross-validation and SHAP-based explanation.	Clinical treatment plan dataset with radiation oncology parameters.	Logistic Regression achieved 93.8 ± 4.1% accuracy and AUC = 0.98 ± 0.02.	Limited to specific clinical dataset; future work can expand model generalization and include additional clinical parameters.
Qjidaa et al. (2020)	Developed a DL-based model using the VGG CNN for image-based clinical decision support. The model included preprocessing, data augmentation, feature extraction, learning, and classification steps.	Medical image dataset for internal and external validation.	Accuracy: 92.5% (internal), 87.5% (external); AUC: 97% (internal), 95% (external); Sensitivity: 92% (internal), 87% (external).	External validation accuracy dropped, suggesting a need for larger and more diverse datasets for robustness.
Hamdaoui et al. (2020)	Presented a clinical decision support system for the prediction of heart illness utilizing ML algorithms on the UCI dataset, including NB, KNN, SVM, RF, and Decision Tree.	UCI Heart Disease dataset containing clinical risk factor attributes.	Naïve Bayes accuracy: 82.17% (cross-validation), 84.28% (train-test split).	Future work suggested applying multi-validation on prospectively collected data for better model approval.
Yahyaoui et al. (2019)	Compared conventional ML models (SVM, RF) with a DL CNN	Pima Indians Diabetes dataset (768 samples, 8 features).	Accuracy: DL (76.81%), SVM (65.38%), RF (83.67%).	Future improvements could involve using hybrid models and larger datasets to enhance DL performance.
Andrianto et al. (2019)	Developed a CDSS for typhoid fever diagnosis using NB, KNN, and SVM classifiers based on medical records.	Medical dataset from Indonesian hospitals related to TFD cases.	Accuracy: KNN (88.7%), higher than NB and SVM.	Further work could integrate more diseases and larger sample sizes for improved diagnostic accuracy.

RESEARCH METHODOLOGY

The methodology of this study involves using the MIMIC-IV dataset as the primary source for evaluating ML models in Clinical Decision Support Systems (CDSS). The data was pre-processed to ensure its quality and consistency by doing things

like resolving missing values, extracting features, and normalizing. In order to assess the model, the data was further partitioned into two parts: training (80%) and testing (20%). They proposed a hybrid XGBoost + Decision Tree (DT) model that can be used to improve the accuracy and generalization of the predictive model using both decision trees and the boosting properties of XGBoost. The metrics that were used to assess performance included accuracy, precision, recall, F1-score and ROC analysis to provide strong and interpretable clinical decision-making predictions. The Multi-Modal Clinical Decision Support System has a proposed workflow, shown in Figure 1.

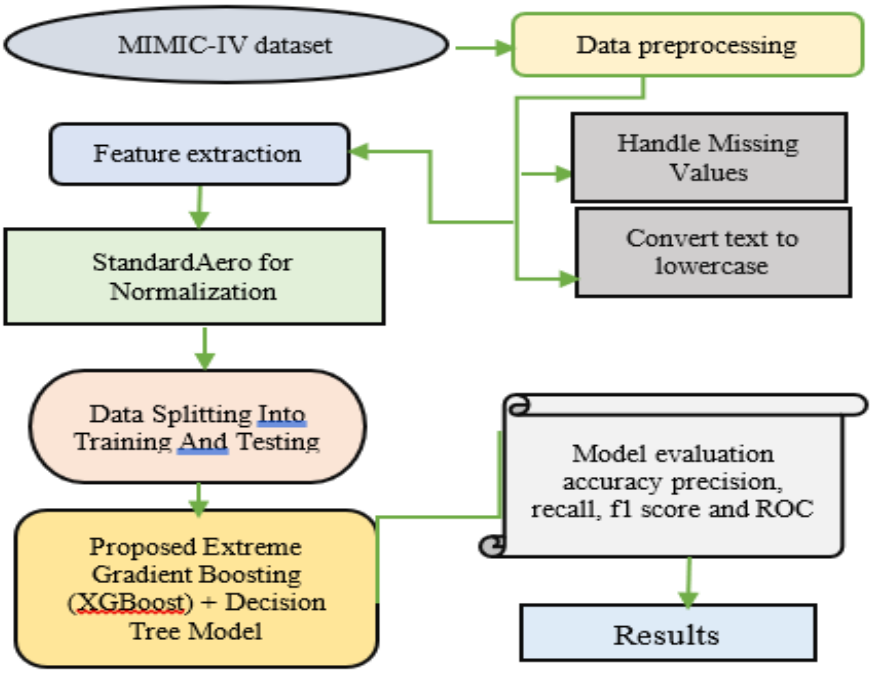

Figure 1 : Proposed Flowchart for Clinical Decision Support Systems

The following section gives a detailed description of each step in the proposed methodology:

A. Data Gathering and Analysis

This paper draws on the Medical Information Mart for Intensive Care IV (MIMIC-IV) data to provide the main data source to assess ML models and explain ability methods in the Clinical Decision Support System (CDSS). It offers a high-quality and multi-modal dataset containing more than 364,000 distinct and unique patient records. Data visualizations such as bar plots and heatmaps were used to examine distribution, feature correlations etc., are given below:

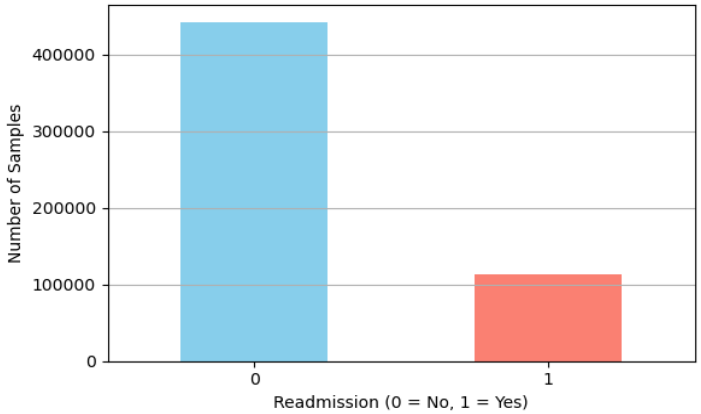

Figure 2: Bar graph of Class Distribution of the Dataset

A notable disparity in the dataset's classes is shown in Figure 2. The majority of samples, represented by bar 0 (indicating No Readmission), are well over 400,000. Conversely, the number of samples for bar 1 (indicating Yes Readmission) is substantially smaller, appearing to be just over 100,000. This demonstrates that the event of an ICU readmission within 30 days is a minority class in this dataset, with roughly four times as many cases of no readmission as cases of readmission.

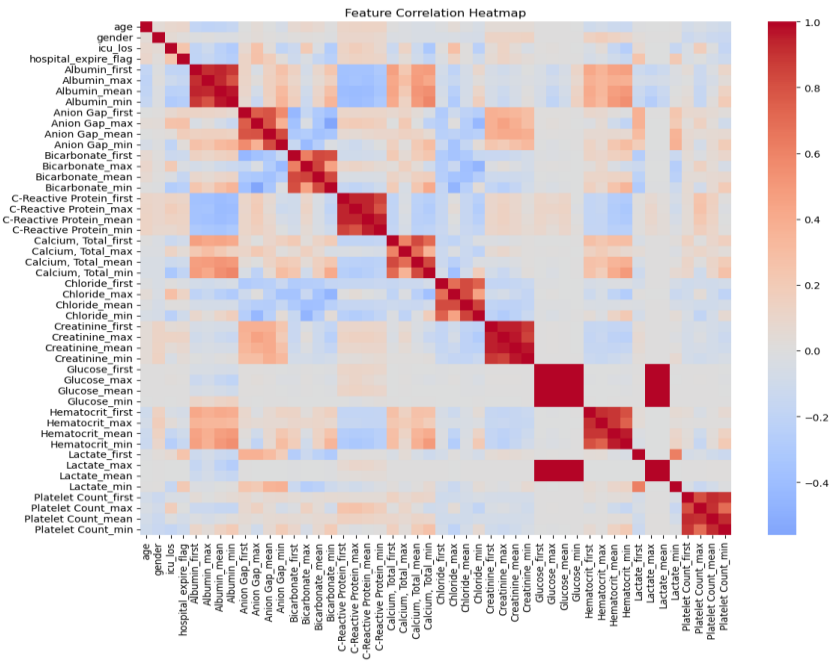

Figure 3 : Correlation Matrix Heatmap of Clinical features

Figure 3 visualizes the pairwise relationships between various clinical and laboratory features. The correlation coefficient between two variables is represented by each cell, and its value can range from -1 to 1. If one characteristic is represented in deep red, it means that For traits A and B, a strong positive association is evident; for features C and D, a strong negative correlation is shown by the color deep blue. The diagonal features are perfectly correlated with themselves (value of 1). Demographic data, such as age and gender, vital signs, and lab values, such as albumin, glucose, creatinine, and platelet counts are included in the heatmap, which indicates patterns of co-variation which can be used to inform feature selection, dimensionality reduction, or predictive modelling in clinical data.

B. Data Pre-Processing

The MIMIC dataset was used in the preparation of the data which consisted of concatenation, cleaning and feature engineering. The pre-processing involved the treatment of missing values, extraction of features normalization. The major pre-processing processes are highlighted below:

- Handle Missing Values: Missing values were handled by setting all correlated events to zero. This method prevents imputation bias while preserving the input matrix's structural integrity.
- Convert text to lowercase: The whole text is written in small letters. This normalization process make words such as "Lung" and "lung" be considered as entities whereby the redundancy of language usage is minimal and enhanced uniformity is made between the dataset.

C. Feature Extraction

Feature extraction, the study used a DenseNet121 convolutional neural network pretrained on ChestX-ray to encode chest X-ray images into 1024-dimensional vectors via global mean pooling. These image embeddings were then used as input to language models or for similarity-based retrieval. For text, reports were tokenized without stemming or lemmatization, and vocabulary sizes were 1,933 for the IU dataset and 12,706 for MIMIC-CXR.

D. Normalization using the StandardScale Mllethod

Considering the varying scales of the descriptive variables, the dataset was transformed using the StandardAero() function to produce a distribution with an average of 0 and a dispersion of 1. As a consequence, the data set was considered standardized. To accomplish this change, we will divide the sum of all the observations by their standard deviation, as indicated in Equation (1):

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

where z represents the feature's modified value, x stands for the initial value of each descriptor, where μ represents the mean of the feature, and σ stands for its standard deviation within the dataset.

E. Data Splitting

The dataset was split using the split function into a training set and a test set so that the classification model could be evaluated. The procedure transmitted 80% of the data to the training set and 20% to the test set. To ensure that training and evaluation samples were representative, the split was stratified to maintain the class proportions in both sets.

F. Proposed Extreme Gradient Boosting (XGBoost) + Decision Tree Model

An ensemble hybrid model of ML, the XGBoost + Decision Tree, is well-liked for structured data categorization and regression in applications like clinical decision support systems. It integrates the interpretability of decision trees (DTs) with the boosting capabilities of XGBoost, which trains one or more learners sequentially to minimize the error in a prediction. Here, every DT is a weak learner, and XGBoost also optimizes a loss function by adding new trees, which concentrate on the error of the older trees. This approach outperforms using a single decision tree in terms of accuracy and generalizability, so it is appropriate for heterogeneous clinical data, including lab tests, vitals and encoded features in other modalities. The XGBoost model predicts the output as the sum of predictions from all trees in Equation. (2)

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F \qquad (2)$$

In the XGBoost + Decision Tree model, the ensemble's anticipated value, $\hat{y}_i$, is calculated by adding the outputs of all the individual decision trees. Here, K represents the sum of all trees included in the model, and where each f_k represents the k-th decision tree. This one is an ineffective learner that attempts to fix the errors caused by the trees that came before it. The set F is the space of all possible regression trees that can be used by the model in training.

At training, XGBoost + DT minimizes an objective that is made up of the training loss and a regularization term to avoid overfitting. The training is performed sequentially, with each subsequent tree fitting the residuals (errors) of the last ensemble. This enables the model to concentrate on the hard-to-predict samples. Regularization also involves tree depth, L1/L2 penalties, and learning rate adjustments, which make XGBoost robust and well-suited to high-dimensional clinical data. Moreover, feature importance scores can be obtained, based on the trained ensemble, and so can be interpreted, which is essential in clinical decision-making. The XGBoost maximizes the objective given by Equation. 3:

$$L(\emptyset) = \sum_{i=1}^{n} l(y_i \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (3)$$

The notation $l(y_i \hat{y}_i)$ is used to denote a differentiable loss (e.g., squared error loss or logistic loss). This regularization and combination of boosted trees enable XGBoost + DT to make accurate and robust predictions using structured clinical data and hence find extensive use in multi-modal clinical decision support systems.

G. Evaluation Metrics

A number of assessment objectives were used to gauge how well the suggested model performed. The classification results were summarized to create a confusion matrix that represented the total number of right and wrong predictions made by each class. Out of this matrix, have obtained values that represent some of the important parameters in the game like True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) that were in turn used to derive important measures of key performance, i.e., accuracy, precision, recall, and F1-score as calculated in Equations (4) to (7):

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+Fp+TN+FN}} \qquad (4)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad (5)$$

$$Recall = \frac{\mathrm{TP}}{TP+FN} \qquad (6)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (7)$$

A trained model's accuracy is the proportion of its accurate predictions to the whole dataset occurrences. As a proportion of all positive occurrences anticipated, precision measures how many accurate positive examples have been identified. Recall measures how many true positives out of a total number of true positives are properly detected. The F1-score strikes a balance by harmonically averaging accuracy and recall. At various decision thresholds, An illustration of the compromise between the true positive and false positive rates is provided by the ROC curve.

IV. RESULTS AND DISCUSSION

The tests were executed on a robust workstation with three Intel CPUs and 64 GB of RAM. Table II shows that the proposed hybrid method, which combines Extreme Gradient Boosting (XGBoost) with Decision Trees (DT), has a 92% accuracy rate in predicting the fate of most cases. Its precision of 96% indicates a high proportion of correctly recognized positive cases, whereas its recall of 95% indicates that the model is efficient at recognizing true positives. The F1-score of 95%

is also another evidence of the balanced precision and recall and shows the reliability and strength of the model in use in clinical decision-making.

Table 2 : Results of the Proposed Model for Clinical Decision Support Systems on the Mimic Dataset

Performance Matrix	Hybrid XGBoost + DT
Accuracy	92
Precision	96
Recall	95
F1-score	95

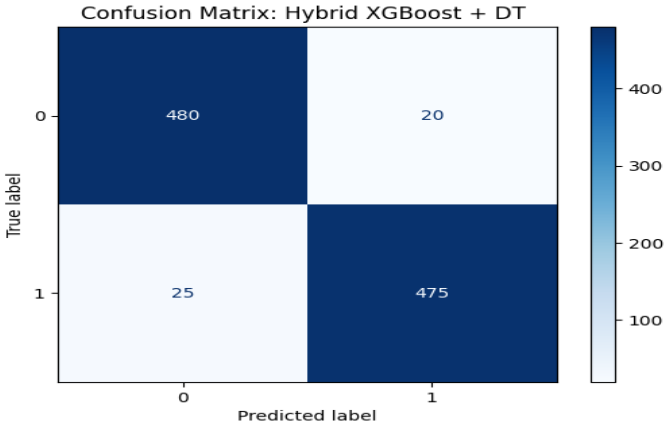

Figure 4 : Confusion Matrix of the Hybrid XGBoost+DT Model

Figure 4 proves it has a great capacity to distinguish between the two classes. It indicates that the model was correct for a large proportion of samples in both categories, with 480 correctly classified as class 0 and 475 as class 1. A limited number of samples were inaccurately reported, with 20 cases of false positives and 25 false negatives, therefore, demonstrating that the model performed well without causing many mistakes. Generally, the matrix underscores the performance and the consistency of the model to predict the occurrence of both good and bad outcomes.

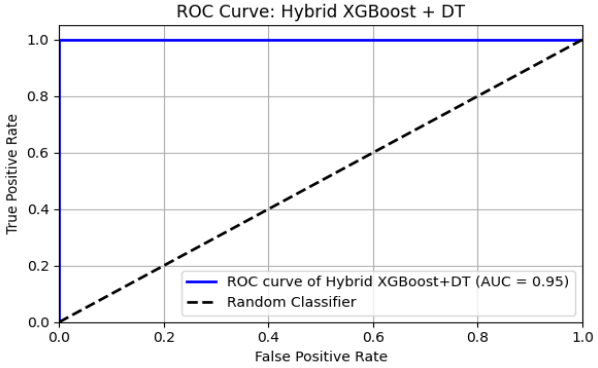

Figure 5 : ROC Curve for XGBoost+DT Model

The classification performance of the proposed Hybrid, XGBoost + Decision Tree (DT) model is shown in Figure 5. The graph shows that the true positive rate and false positive rate are well separated, indicating that the classification between classes is successful. Being AUC=0.95, the model has very good predictive capacity, with a high sensitivity and specificity. The model's ability to minimize false positives and maximize genuine positives is demonstrated by the curve's proximity to the top left side, which proves its strength and effectiveness in the classification of medical data.

A. Comparative Analysis

A comparative evaluation of several ML and DL systems used as clinical decision support systems depicts significant variations in predictive results in Table III. The traditional ML approach using Random Forest (RF) achieved 77% accuracy, indicating its ability to work with structured clinical data, although it had limited capacity to capture complex patterns. DL models showed a greater performance with LSTM networks scoring 82.01% and Text-Image Embedding Network (TieNet) scoring 84.8% as they processed textual and image features to enhance clinical representation. The developed hybrid model

was more effective than all other methods. The given comparison highlights the benefits of mixed methods in improving decision support in the healthcare context.

Table 3 : Comparison of Different Machine Learning and Deep Learning Models for Clinical Decision Support Systems

Model	Accuracy
Random Forest (RF) [18]	77
LSTM (Long Short-Term Memory) [19]	82.01
TieNet (Text-Image Embedding Network)[20]	84.8
Proposed model	92

The suggested Hybrid XGBoost + Decision Tree model is rather beneficial as it provides the interpretability of a decision model with the predictive capabilities of gradient boosting. It improves the generalization, stability, and robustness, which is why it is very appropriate for the complicated clinical setting. In comparison to conventional or standalone models of DL, the hybrid one successfully represents both linear and non-linear relationships with various clinical data. Its incorporation of multi-modal enhances further the decision support, where it makes accurate, explainable, and reliable predictions that can be used to guide healthcare professionals in making informed clinical decisions.

V. CONCLUSION AND FUTURE STUDY

The new technology, electronic health records, might result in more data that is constantly being taken about individual patients and more information that can be reviewed about previously treated patients. Nevertheless, so far, such data has not been used effectively to develop the clinical decision support systems that will be used to generate personal clinical advice to help clinicians deliver personalized healthcare. According to the outcomes of the experiment, the comparison of different foundation models of Multi-Modal Clinical Decision Support Systems shows that traditional ML models, such as Random Forest (77%), DL models, such as LSTM (82.01%) and Tie Net (84.8%), are also effective in clinical prediction. But, the suggested hybrid XGBoost + Decision Tree model has the highest accuracy of 92%, which shows its better ability to identify the intricate patterns within multi-modal clinical data. The importance of this major enhancement is that the hybrid ensemble strategy effectively improves dependability, interpretability, and diagnostic accuracy of clinical decision support applications. future research investigations should focus on user verification, time series modeling, and more comprehensive comparison analyses of explain ability methods in different healthcare environments.

VI. REFERENCES

[1] J. Bajwa, U. Munir, A. Nori, and B. Williams, “Artificial intelligence in healthcare: transforming the practice of medicine,” Futur. Healthc. J., vol. 8, no. 2, Jul. 2021, doi: 10.7861/fhj.2021-0095.

[2] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, “Secure and Robust Machine Learning for Healthcare: A Survey,” IEEE Rev. Biomed. Eng., vol. 14, pp. 156–180, 2021, doi: 10.1109/RBME.2020.3013489.

[3] T. K. Mackey et al., “‘Fit-for-purpose?’ – challenges and opportunities for applications of blockchain technology in the future of healthcare,” BMC Med., vol. 17, no. 1, p. 68, Dec. 2019, doi: 10.1186/s12916-019-1296-7.

[4] A. Papa, M. Mital, P. Pisano, and M. Del Giudice, “E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation,” Technol. Forecast. Soc. Change, vol. 153, p. 119226, Apr. 2020, doi: 10.1016/j.techfore.2018.02.018.

[5] M. Carlile, B. Hurt, A. Hsiao, M. Hogarth, C. A. Longhurst, and C. Dameff, “Deployment of artificial intelligence for radiographic diagnosis of COVID-19 pneumonia in the emergency department,” JACEP Open, 2020, doi: 10.1002/emp2.12297.

[6] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, “TieNet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-Rays,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018. doi: 10.1109/CVPR.2018.00943.

[7] B. Jing, P. Xie, and E. P. Xing, “On the automatic generation of medical imaging reports,” in ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 2018. doi: 10.18653/v1/p18-1240.

[8] A. Brady, R. Ó. Laoide, P. McCarthy, and R. McDermott, “Discrepancy and error in radiology: Concepts, causes and consequences,” 2012.

[9] M. Ranganathan et al., “Attenuation of ketamine-induced impairment in verbal learning and memory in healthy volunteers by the AMPA receptor potentiator PF-04958242,” Mol. Psychiatry, vol. 22, no. 11, pp. 1633–1640, Nov. 2017, doi: 10.1038/mp.2017.6.

[10] S. Gössling, “Risks, resilience, and pathways to sustainable aviation: A COVID-19 perspective,” J. Air Transp. Manag., vol. 89, p. 101933, Oct. 2020, doi: 10.1016/j.jairtraman.2020.101933.

[11] R. Rybnicek, J. Plakolm, and L. Baumgartner, “Risks in Public–Private Partnerships: A Systematic Literature Review of Risk Factors, Their Impact and Risk Mitigation Strategies,” Public Perform. Manag. Rev., 2020, doi: 10.1080/15309576.2020.1741406.

[12] S. V. B. Jardim, “The Electronic Health Record and its Contribution to Healthcare Information Systems Interoperability,” Procedia Technol., vol. 9, pp. 940–948, 2013, doi: 10.1016/j.protcy.2013.12.105.

[13] P. Siciarz, S. Alfaifi, E. Van Uytven, S. Rathod, R. Koul, and B. McCurdy, “Machine learning for dose-volume histogram based clinical decision-making support system in radiation therapy plans for brain tumors,” Clin. Transl. Radiat. Oncol., vol. 31, pp. 50–57, Nov. 2021, doi: 10.1016/j.ctro.2021.09.001.

[14] M. Qjidaa et al., "Development of a clinical decision support system for the early detection of COVID-19 using deep learning based on chest radiographic images," in 2020 International Conference on Intelligent Systems and Computer Vision, ISCV 2020, 2020. doi: 10.1109/ISCV49265.2020.9204282.

[15] H. El Hamdaoui, S. Boujraf, N. E. H. Chaoui, and M. Maaroufi, "A Clinical support system for Prediction of Heart Disease using Machine Learning Techniques," in 2020 International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2020, 2020. doi: 10.1109/ATSIP49331.2020.9231760.

[16] A. Yahyaoui, A. Jamil, J. Rasheed, and M. Yesiltepe, "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," in 2019 1st International Informatics and Software Engineering Conference (UBMYK), IEEE, Nov. 2019, pp. 1–4. doi: 10.1109/UBMYK48245.2019.8965556.

[17] B. Andrianto, Y. K. Suprapto, I. Pratomo, and I. Irawati, "Clinical decision support system for typhoid fever disease using classification techniques," in Proceedings - 2019 International Seminar on Intelligent Technology and Its Application, ISITIA 2019, 2019. doi: 10.1109/ISITIA.2019.8937286.

[18] W. Boag, T. M. Harry Hsu, M. McDermott, G. Berner, E. Alsentzer, and P. Szolovits, "Baselines for Chest X-Ray Report Generation," Proc. Mach. Learn. Res., vol. 116, pp. 126–140, 2019.

[19] S. P. Tembhurne, "Clinical Decision Making Using Machine Learning and ICU Data," Helix, vol. 8, no. 5, pp. 4082–4087, 2018, doi: 10.29042/2018-4082-4087.

[20] Z. Babar, T. van Laarhoven, and E. Marchiori, "Encoder-decoder models for chest X-ray report generation perform no better than unconditioned baselines," PLoS One, 2021, doi: 10.1371/journal.pone.0259639.