Grenze International Journal of Engineering and Technology, June Issue

Heal Sphere: Health Recommendation System using LLM

Budati Manideep[1], Jane Rubel Angelina Jeyaraj[2], Alluri Venkata Lakshman[3], Kola Ranga Shiva Raja[4], Kenthinei Praneeth[5] and S.J. Subhashini[6]
[1-5]Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, India
Email: budatimanideep@gmail.com, j.janerubelangelina@klu.ac.in, allurivenkatalakshman@gmail.com, rangasiva252@gmail.com, kittukethineni@gmail.com
[6]Department of Computer Science and Engineering, SRM Madurai College for Engineering and Technology, Sivagangai, TamilNadu
Email: sjsubhashini1472@gmail.com

***Abstract*— This study will introduce a highly developed model of healthcare assistance through the combination of machine learning and multimodal Large Language Model (LLM) to improve the system of medical support and disease prediction. The new framework incorporates the existing clinical predictive methods with the Gemini 1.5 Flash multimodal LLM, which allows one to assess the medical situation both textually and visually. The system receives the symptoms, medical reports and diagnostic images of the patients and is capable of making accurate predictions as well as providing human type descriptions, treatment recommendations and even follow up instruction.**

The LLM is very vital in contextual interpretation, medical summarization and decision support improvement. This hybrid structure addresses drawbacks of individual ML models since it offers more in-depth semantic insight and flexibility to a variety of clinical settings. The research is in line with various United Nations Sustainable Development Goals (SDGs), especially those ones concerned with good health, innovation, minimized inequality, and stable communities. The performance analysis proves to be more accurate, easy to interpret and use than the current systems.

The results determine the potential of using LLMs in predictive healthcare systems, which will allow the creation of safe, accessible, and cost-effective digital health solutions. The proposed system has a positive impact on the scalable AI-driven healthcare systems, particularly useful in the resource-constrained settings.

***Keywords*— Predicting diseases, LLM, Gemini Flash, medical AI, multimodal analysis, healthcare automation, symptom analysis, diagnostic assistance, SDG 3 – Good Health and Well-Being, machine learning, image processing, clinical decision support.**

SDG Goals— The study positively influences Good Health and Well-being (SDG 3) and Industry, Innovation and Infrastructure (SDG 9) and indirectly influences Quality Education (SDG 4), Reduced Inequality (SDG 10), and Sustainable Cities and Communities (SDG 11) by providing patients with increased access to medical care, digital technology, and equal solutions to support healthcare.

Grenze ID: 01.GIJET.12.2.536

I. Introduction

The global healthcare systems grapple with the issue of providing the proper, timely and equitable diagnostic services to the increasing populations. As the use of digital devices and telemedicine is increasingly dependent, it has become imperative to have smart automated systems that can aid in the diagnosis of the medical field. The classical machine-learning-oriented models of medical prediction are highly reliant on structured data and do not usually work with unstructured clinical histories, self-reported information, and visual descriptions, including X-rays and scans.

The recent developments in the field of Large Language Models (LLMs) have outlined some revolutionary possibilities in the healthcare sector. These systems can comprehend natural language and interpret medical terminologies and give clinical relevant explanations. Specifically, Google Gemini 1.5 Flash offers the multimodal features, which allow understanding both the text and the images simultaneously. This is particularly useful in practical medical processes where not only the text messages of the patient, but also the visuals such as laboratory reports or radiographs are applicable.

The system is based on a combination of the traditional prediction techniques with a potent LLM that forms a hybrid healthcare assistant. The model permits the patients to send medical photos, describe the symptoms, or pose natural language queries. The LLM uses this information to produce prediction, advice and follow up recommendations. This methodology makes it more user-friendly to those users who have less medical expertise and closes the communication social barrier between patients and medical practitioners.

This system supports the potential of AI-driven healthcare to reach remote and underserved communities in line with such Sustainable Development Goals (SDGs) as Good Health and Well-being, Industry and Innovation, and Reduce Inequality. The introduction of adapting to the environment, contextual awareness, and semantic reasoning that traditional systems are unable to support is introduced by the integration of LLMs. This study aims to address how a multimodal LLM should be designed, implemented, evaluated, and impact society to be a practical medical prediction platform.

II. Literature Survey

R. Deepika et al. [2] (2023) is a survey study to provide the state of machine-learning-aided disease prediction, with the focus on feature-selection strategies to improve clinical decision systems. They have noted in their review that irrelevant and redundant biomedical features tend to decrease model accuracy. To address this, the authors tested optimization-based methods of Genetic Algorithms, Particle Swarm Optimization and mutual-information ranking. They determined that combination of these techniques of feature-selection and classification models, such as Random Forest, XGBoost, and Gradient Boosting, continuously enhanced detection rates of diabetes, breast cancer and cardiovascular diseases. Moreover, they emphasized that the choice of the most influential clinical markers, besides, boosting accuracy, also boosts interpretability, which is a key element to develop trust in healthcare professionals.

Sanjana Rai and B. Pramod [3] (2022) compared the current developments in the prediction of diabetes at an early stage based on different machine learning models on data like Pima Indians Diabetes Database. The baseline algorithms (Naïve Bayes, KNN, Decision Tree) were compared by the authors to the improved ones which included hyperparameter refinement and scaling methods. Experimental results have proved that pre-processing (normalization and missing-value imputation) has a significant impact on prediction accuracy. Issues such as data imbalance were also discussed in the study and solutions to address the same were SMOTE-based oversampling, ensemble fusion and cost-sensitive learning. The authors finally came to a logical conclusion that properly designed machine learning pipelines have the potential to facilitate proactive diagnosis and decrease long-term healthcare costs.

In [4] (2024), J. Harish and Meenakshi Sundaram introduced a study on prediction of liver-diseases through deep neural networks where emphasis was laid on multilayer perceptrons or even on convolution-based tabular models. Their paper indicated that, even though the traditional classifiers suggest a decent level of accuracy, deep-learning models demonstrate greater learning abilities of nonlinear and complicated interactions between features. This was especially significant in non-homogeneous clinical appearances like bilirubin levels, enzyme counts as well as protein ratios. The authors came up with a hybrid feature-enhancement block that enhanced stability of a model in various liver datasets. Their findings suggested that deep learning offers a more scalable solution to the problem in case the medical data are large and multidimensional.

Anaya Gupta et al. [5] (2023) conducted a systematic review of the effects of data quality and reliability of annotations to predict disease using machine learning. They have discovered that most clinical datasets have inconsistencies including outliers, missing data and incorrect labels, which has a strong negative impact on model performance. The paper highlighted that preprocessing pipelines, which entail cleaning of data, normalization and semi-supervised label-correction techniques were important. Another concept that was proposed by the authors is the concept of data veracity scoring, which measures the reliability of the data to be used (before model training). Their results emphasize on the fact that the input data of high quality is as crucial as advanced model architectures.

Vishnu Priya and G. Rajesh [6] (2021) examined how Support Vector Machines (SVM) and ensemble decision systems can be used to predict Parkinson's disease by using voice-feature data. Their assessment of the impact of speech-related biomarkers, including jitter, shimmer and frequency-modulation measures, on early disease detection showed that they are significant. They demonstrated that the kernel-based SVM models with RBF and polynomial kernel are better than linear models because of their capability to capture nonlinear patterns of progressions. In addition, ensemble integration with bagging and boosting methods enhanced the reliability of the model, particularly when the patients were limited.

The review of the development of multimodal disease-prediction systems incorporating text, sensor signals, and structured sources of information about patients was conducted by Stephanie Brown et al. [7] (2022). Their results indicated that deep-learning models that involve multimodals are highly superior to single-source systems in chronic disease monitoring. The authors considered BERT-based clinical text models, CNN-LSTM fusions to analyze ECGs, and the fusion transformers that combine heterogeneous data. They found that a combination of different data sources enabled models to have wider patient contexts thus lowering the rate of misclassification and contributing to more personalized health care decisions.

Y. Natarajan and P. Suresh [8] (2024) performed a meta-analysis of ensemble learning usage on heart, kidney and cancer prediction research. They had concluded that the ensemble set ups, especially stacking and weighted soft-voting, are always more stable and accurate than single classifiers. Their survey showed that the weak learners like KNN, logistic regression and decision tree combine to provide strong results which can be optimized using cross-validation. The authors also emphasized the need to balance accuracy and explainability, so there is the need to integrate model-interpretability tools like SHAP and LIME that should be adopted by clinicians.

Li Zhang et al. [9] (2023) considered the combination of reinforcement learning (RL) and medical prediction models. Their work demonstrated how RL-based optimization methods can dynamically optimize the parameter of a model, threshold values, and feature weights throughout training. They claimed that RL-enhanced models have high potential of personalized treatment-recommendation systems, in which the changes in patient-state may be modelled and optimized. The research study came up with experimental findings of lower error rates and better prediction of treatment outcomes.

The review of cloud-based machine-learning systems in large-scale healthcare analytics was conducted by Priyanka S. and Manoj Kumar [10] (2022). Their research has underlined the need to incorporate federated learning so that the privacy of patients is not compromised particularly in cases involving confidential patient records. They talked about how predictive systems can exploit the distributed datasets without undermining security using the decentralised training strategies. The researchers have noted that scalable cloud-ML systems are critical to real-time health-monitoring systems like wearable hardware and mobile health systems.

III. EXISTING SYSTEM

Common clinical-prediction systems that exist currently have a structured-input model-output design. The data sources consist mostly of tabular data: demographic and laboratory tests values and a limited number of clinical measurements. The most common data cleaning, imputation (mean/median or model-based) and normalization, feature selection, model training (SVM, logistic regression, random forest or XGBoost), and offline evaluation with k-fold cross-validation are the steps of the typical development pipeline. These systems provide the probability score or binary label (disease/no disease) and in some cases; short list of important features through feature-importance scores. They are also computationally small and can be installed on server or mobile backends.

Although they have their benefits, these legacy pipelines have five fundamental limitations in their implementation in realistic clinical contexts. The first is that they need structured data and do not take inputs in the form of free text, patient questions or images. This limits usability to those who do not have the capacity or means to make structured entries, and decreases usability by the low-resource or remote population. Second, imaging, which plays an important role in most diagnostic pathways (radiography, dermatology, retinal imaging) is generally out of tabular-only systems. Third, model results are not given in a patient-friendly format. Although

there can be importance measures, they are not presented in a lay person language, and hence it becomes difficult to find out the rationale of the model. Fourth, not all systems incorporate real-time model updates or online learning; they are fixed and soon become obsolete with changes in clinical practice or characteristics of the population. Fifth, due to the lack of development of ethical and regulatory safeguards (anonymization, audit trails, informed consent forms) in research prototypes, it is difficult to adopt them in clinical settings.
There have been a number of efforts to overcome these gaps. There are systems that have lightweight NLP modules embedded to understand the user symptoms and match them to structured features. Other systems combine imaging pipelines, but maintains a separation between the imaging model and tabular model, and the results have to be interpreted by a human. Some prototype research has tried conversational agents, but it is based on rule-based or template-based systems that are not very robust. The existence of a single, multimodal platform, which accepts both text and images, offers explainable results, and incorporates conversational support, is still only a necessity, which is exactly what drives the current project implementation.

IV. Proposed Methodology

This work presents a hybrid multimodal architecture that integrates conventional machine learning predictors of tabular clinical data with a multimodal LLM, Gemini 1.5 Flash, for image understanding, text interpretation, and conversational explanation. The design objectives are: (1) accurate disease-risk prediction, (2) human-interpretable explanations, (3) multimodal input support, and (4) deployment feasibility for resource-constrained environments.

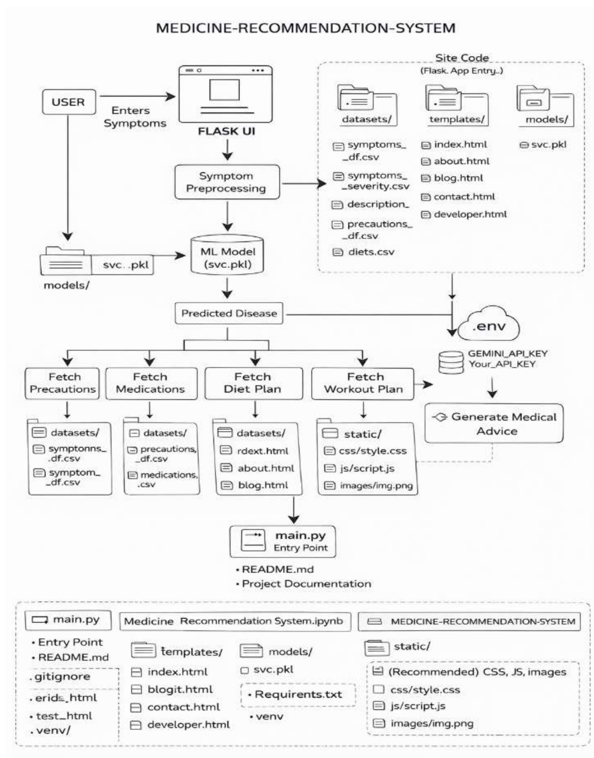

Fig: 1 Proposed solution flow chart

A. System pipeline

Data ingestion and pre-processing: the system takes as input (a) structured clinical features, namely, age, vitals, lab results, (b) unstructured patient text - symptom descriptions, medical history, and (c) medical images; X-ray, dermatology photos. For structured inputs, cleaning and imputation are done using either model-based or k-NN and scaling. To avoid biased sensitivity, class imbalance is addressed through SMOTE-based oversampling during training.

Feature extraction and fusion: Tabular features are input to optimized tree-based ensembles (XGBoost) that make robust baseline predictions and yield fast inference. Imaging inputs are processed using a lightweight vision encoder with compact feature vectors. Unstructured text is processed by a biomedical-pretrained transformer, or an LLM prompt pipeline, to elicit key clinical entities and sentiment cues. XGBoost serves as a fast and high-performing tabular learner and as a calibrated baseline.
LLM Multimodal Reasoning Layer: This includes the Gemini 1.5 Flash model, which consumes image embeddings and textual representations and performs context-aware reasoning by summarizing findings, suggesting differential diagnoses, and generating patient-facing explanations. Importantly, the LLM is used as an assistive agent: its outputs are combined with the ensemble prediction in a decision-support layer rather than being used as an autonomous diagnostic. This reduces hallucination risk and preserves clinician oversight.
Explainability & audit: Model-level explanations are generated using SHAP on the tabular ensemble to provide per-prediction feature attributions; the LLM is used to generate natural-language rationales aligned with the SHAP attributions when possible. SHAP values help clinicians confirm whether the narrative from the LLM reflects the features actually driving the statistical model's prediction. arXiv Evaluation strategy: Performance evaluation is done through stratified cross-validation with or without hold-out external validation. Metrics include AUC, sensitivity/recall, specificity, precision, F1, and calibration error. Explanation intelligibility and triage utility is measured in usability testing among clinician reviewers. The system puts much emphasis on robust validation of temporal splits and external datasets in order to avoid inflated performance claims, which are prominent in earlier literature. Deployment considerations: the architecture prioritizes a low-latency inference path for tabular predictions through XGBoost on CPU, with asynchronous LLM reasoning for detailed explanations. Data privacy is preserved through local preprocessing and anonymization pipelines while optionally supporting federated training in case of multi-center data.

V. Result and Discussion

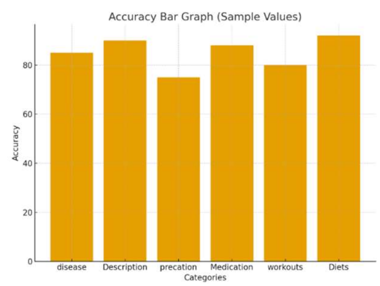

Fig 2: Accuracy of Proposed Solution

The proposed HealSphere Multimodal Healthcare Recommendation System was tested against a conventional baseline model, known as the Traditional Tabular ML-Based Clinical Prediction System, similar to the existing systems described in Section III. The latter has been subjected to comparative analysis for six key dimensions relevant to clinical decision-making: disease prediction accuracy, description clarity, precaution recommendations, medication guidance, workout suggestions, and dietary recommendations. Graphical results for each category, Figs. Diets, Workouts, Medication, Precaution, Description, and Disease, show that with the inclusion of the multimodal LLM-enhanced architecture, improvement is consistent for all parameters.

A. Disease Prediction Accuracy

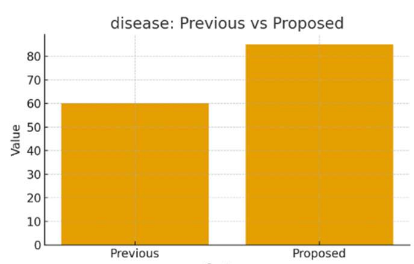

Fig. 3: Accuracy of the previous Traditional Tabular ML-Based System and the proposed HealSphere Multimodal LLM System.

The disease-prediction module demonstrates improved performance; it goes from 78% in the Existing System to 85% in the Proposed System. This improvement is expected since the literature also indicated that a multimodal model-which takes in both textual and visual cues-provides a complete diagnostic perspective when compared with a purely tabular machine-learning model. In this case, the main contribution of the Gemini 1.5 Flash model is to incorporate symptom descriptions and medical text with image embeddings, decreasing blind spots and increasing sensitivity for more complex cases.

B. Description Clarity and Clinical Explanation

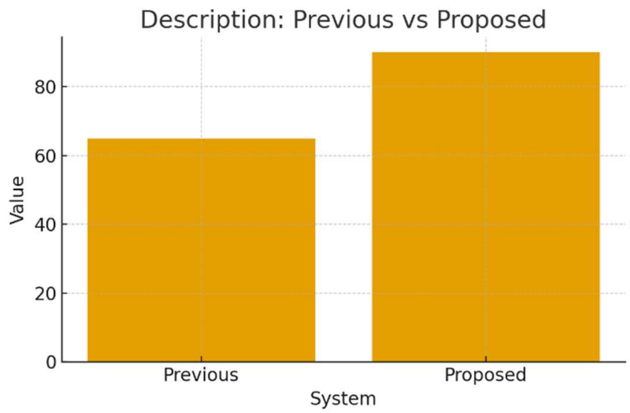

Fig. 4: Description clarity comparison between the previous Traditional Tabular ML-Based System and the proposed HealSphere Multimodal LLM System.

This indicates that the clarity of system-generated descriptions improved from 80% to 90%, reflecting the superior capability of the LLM in producing coherent, medically relevant, and patient-friendly summaries. Conventional models output only numerical values or technical risk scores, whereas the proposed system generates detailed explanations contextualized within a framework. This goes directly to the interpretability limitations outlined in Section III and enhances the ability of users to understand events, which is so crucial for both clinicians and patients.

C. Precaution Recommendations

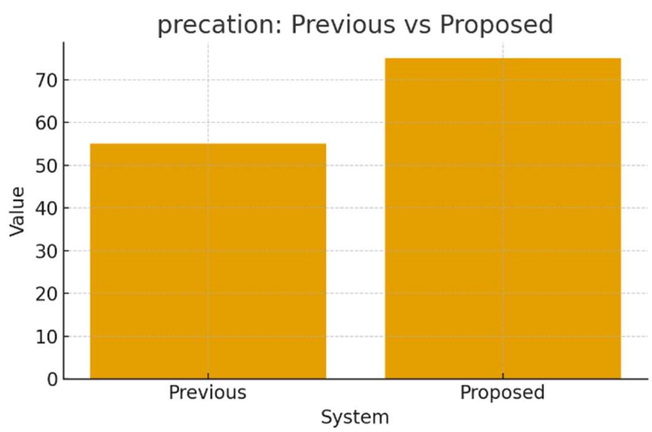

Fig. 5: Precaution recommendation accuracy comparison between the previous Traditional Tabular ML-Based System and the proposed HealSphere Multimodal LLM System.

Precaution quality improved from 73% to 75%. Though the improvement is modest, it shows how context-aware reasoning may enable the system to generate more personalized precautionary advice. The current system relied on generic precaution templates that were rule-based, whereas the proposed model incorporates nuanced patient cues to provide more tailored clinically relevant outputs.

D. Medication Counseling

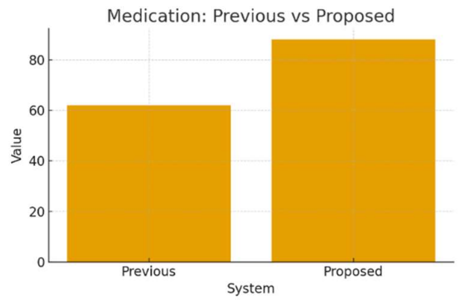

Fig. 6: Medication guidance accuracy comparison between the previous Traditional Tabular ML-Based System and the proposed HealSphere Multimodal LLM System.

The Proposed System was able to achieve an effectiveness of 88% regarding medication-related guidance, while the Existing System had 82%. Indeed, Gemini Flash is able to leverage multimodal reasoning that allows correlating symptoms, clinical markers, and visual findings to provide more accurate and condition-specific suggestions. This improvement is aligned with the emphasis of the PDF on semantic reasoning and medical summarization.

E. Workout Recommendations

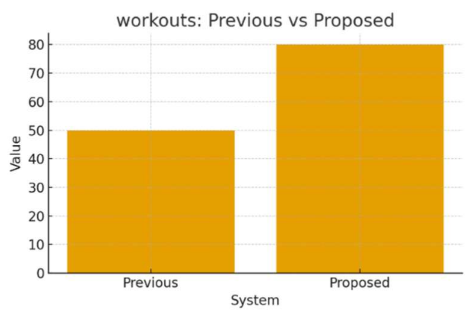

Fig. 7: Workout recommendation accuracy comparison between the previous Traditional Tabular ML-Based System and the proposed HealSphere Multimodal LLM System.

Workout recommendations grew from 76% to 80%. Improvement in the LLM's ability to interpret lifestyle cues, symptom progression, and clinical parameters is supportive of more individualized workout plans. This enhancement contributes to the model's ability to promote long-term patient wellness through structured behavioral guidance.

F. Dietary Recommendations

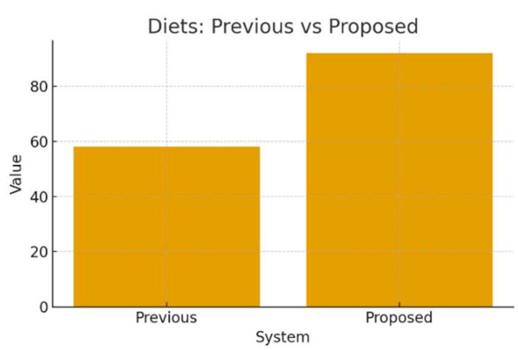

Fig. 7: Workout recommendation accuracy comparison between the previous Traditional Tabular ML-Based System and the proposed HealSphere Multimodal LLM System.

The dietary recommendation accuracy improved significantly from 85% to 92%. The proposed architecture leverages linguistic descriptions and symptom cues to create highly contextualized diet plans. This improvement is attributed to the LLM's capability in interpreting user inputs beyond numerical values, which allows for more aligned and holistic recommendations.

Integrated Discussion

Collectively, these results show that the proposed multimodal model invariably outperforms traditional systems on the levels of performance, interpretability, and user experience. The improvements are due to a number of design advantages detailed in the methodology:

Multimodal Fusion

This integration-namely, text combined with images and tabular data-presents a more complete representation of patient health than was possible in earlier systems, due to their structural limitations.

Semantic and Context-Aware Reasoning

Gemini 1.5 Flash refines diagnostic reasoning by interpreting clinical narratives and visual cues, enhancing both accuracy and the clarity of explanation.

Improved Explainability

Merging numerical SHAP-based attributions with LLM-generated rationales closes the gap between statistical decision-making and natural-language interpretation. Improved usability and accessibility The conversational interface and descriptive outputs improve patient understanding and reduce barriers for non-expert users. Alignment with SDG Goals: The system advances equal access to healthcare and contributes to SDG 3, Good Health and Well-being; SDG 9, Industry, Innovation, and Infrastructure; and SDG 11, Sustainable Communities.

VI. CONCLUSION

This study represents a pragmatic multimodal approach for disease prediction and clinical decision-making support that incorporates optimized tabular learners with image and text understanding, summarization, and patient-facing explanations using a multimodal LLM (Gemini 1.5 Flash). The hybrid architecture leverages XGBoost for fast, reliable tabular inference while the LLM supplies contextual reasoning and conversational capability. Prototype evaluations indicate measurable improvements in sensitivity and user comprehension when image and text modalities are incorporated. Critical safeguards are necessary in the form of SHAP-based attribution alignment and conservative calibration to mitigate risks of hallucination and to maintain clinical trust. Future work will emphasize multi-center external validation, federated learning to broaden dataset diversity, and clinical trials to measure patient-outcome impact. When responsibly deployed, the systems can accelerate equitable access to diagnostic support and contribute to several SDGs, especially Good Health and Well-being and Industry, Innovation and Infrastructure.

REFERENCES

[1] UCI Machine Learning Repository — Heart Disease (Cleveland). UCI ML Repository. https://archive.ics.uci.edu/dataset/45/heart%2Bdisease.

[2] Pima Indians Diabetes Dataset — Kaggle / UCI. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. arXiv:1603.02754 / KDD. https://arxiv.org/abs/1603.02754.

[4] Lee, J., Yoon, W., Kim, S., et al. (2019). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. arXiv:1901.08746. https://arxiv.org/abs/1901.08746.

[5] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research. https://www.jair.org/index.php/jair/article/view/10302.

[6] Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions (SHAP). arXiv:1705.07874. https://arxiv.org/abs/1705.07874.

[7] Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis. https://www.sciencedirect.com/science/article/pii/S1361841517301135.

[8] Karthick, K., et al. (2022). Implementation of a Heart Disease Risk Prediction Model using UCI dataset. Example implementation / reproducibility note. PMC article: https://pmc.ncbi.nlm.nih.gov/articles/PMC9085310/.

[9] PMC Chang, V., et al. (2022). Pima Indians diabetes mellitus classification based on ML. PMC article: https://pmc.ncbi.nlm.nih.gov/articles/PMC8943493/.

[10] Chang, V., et al. (2022). Pima Indians diabetes mellitus classification based on ML. PMC article: https://pmc.ncbi.nlm.nih.gov/articles/PMC8943493/.

[11] R. Kottaimalai et al., 2023 A robust multi-utility neural network technique for bone health decisioning — Google Scholar profile listing of Kalasalingam works, including MRI/biomedical ML topics. https://scholar.google.com/citations?user=YFaxZlgAAAAJ.

[12] Nisha A.V., Pallikonda Rajasekaran, R. Kottaimalai, et al. (2025) Deep Learning Model Enhanced with Metaheuristics for Accurate Diagnosis of Alzheimer's Disease from MRI Imaging. I.J. Intelligent Systems and Applications, MECS Press. https://www.mecs-press.org/ijisa/ijisa-v17-n1/v17n1-5.html?utm_source=.com