Online-Journals.org

International Journal of
Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 11 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i11.56089

PAPER

Intelligent Decision Support System Based on Heart Failure

Ghlana Amara[1,2] (✉),
Abderrahmane Ez-Zahout[1,2],
Soumia Ziti[1,2]

[1]Intelligent Processing & Security of Systems (IPSS) Research Team, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco

[2]Department of Computer and Sciences, Mohamed V University, Rabat, Morocco

ghlana_amara@um5.ac.ma

ABSTRACT

In the ever-changing realm of healthcare, the integration of advanced technologies and the wealth of medical data presents exciting opportunities to enhance diagnostic processes. This study presents the intelligent diagnostic decision support system (IDDSS), employing data analysis and artificial intelligence (AI) to aid doctors in making informed diagnostic decisions. The IDDSS harnesses a wealth of patient data, encompassing medical records, test findings, and demographic details. By employing advanced data analysis techniques, it reveals valuable insights, identifies patterns, and establishes correlations within the data. Furthermore, to elevate diagnostic precision, the IDDSS integrates state-of-the-art AI algorithms and machine learning models. These models, trained on extensive datasets, excel in recognizing intricate patterns, categorizing illnesses, and predicting outcomes. Continuously adapting to incorporate the latest medical advancements, the IDDSS remains at the forefront of enhancing healthcare efficacy and patient care. This system has significant potential to advance healthcare diagnostics, ultimately serving as a valuable decision support tool enabling physicians to provide exceptional care and improve patient outcomes.

KEYWORDS

machine learning, healthcare, artificial intelligence (AI), big data, logistic regression, prediction

1 INTRODUCTION

The rise of big data analytics and artificial intelligence (AI) has transformed many industries, and healthcare is no exception. With the progress of healthcare technology and the abundance of medical data, opportunities to advance diagnostic decision-making have arisen [1], [2]. Our project presents an intelligent diagnostic decision support system (IDDSS) that uses big data analytics and AI to enhance healthcare diagnoses.

The IDDSS acts as a valuable support tool for healthcare professionals, assisting them in making accurate and timely diagnoses.

It utilizes big data analytics [3] and AI algorithms to analyze a vast dataset that includes medical records, clinical measurements, and patient demographics. By uncovering

Amara, G., Ez-Zahout, A., Ziti, S. (2025). Intelligent Decision Support System Based on Heart Failure. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(11), pp. 97–115. https://doi.org/10.3991/ijoe.v21i11.56089

Article submitted 2025-04-25. Revision uploaded 2025-06-18. Final acceptance 2025-06-18.

© 2025 by the authors of this article. Published under CC-BY.

insights, patterns, and connections within the data, the system provides healthcare professionals with a comprehensive understanding of their patients' health, enabling them to make informed diagnostic decisions that can significantly impact patient care [4].

In order to enhance diagnostic precision, the IDDSS uses AI algorithms and machine learning models [5]. These models are trained on large-scale datasets to recognize patterns, classify diseases, and predict patient outcomes. By continuously learning from new data, the system improves its diagnostic capabilities over time, ensuring that it stays current with the latest medical research and findings.

The IDDSS features a user-friendly interface that presents diagnostic recommendations based on the patient's data analysis. Healthcare professionals can easily access and interpret the system's insights, enabling them to make informed decisions and provide personalized treatment plans. The system also promotes knowledge sharing and collaboration among healthcare providers, allowing for collective intelligence in the diagnostic process.

By integrating big data analytics and AI, the IDDSS can revolutionize healthcare diagnoses. It enables the extraction of actionable insights from vast amounts of medical data, leading to more accurate and timely diagnoses. The system serves as a valuable tool for healthcare professionals, empowering them to deliver enhanced healthcare outcomes and improve patient care [6].

In this project, we will develop a predictive model for heart failure [7] using a dataset containing various clinical features [8], [9]. We will perform data preprocessing, exploratory data analysis, and feature engineering to prepare the data for model training. The trained model will be evaluated using various performance metrics [10], and the results will be integrated into the IDDSS interface. The final system will provide healthcare professionals with valuable diagnostic support, enabling them to make informed decisions and improve patient outcomes.

2 RELATED WORK

Several recent studies published have explored the use of machine learning for cardiovascular and heart failure prediction:

- Logistic regression combined with grid search was applied to clinical data for heart disease prediction, showing how hyperparameter tuning can improve classification accuracy [11].
- Another research in Artificial Neural Networks with K-Fold Cross-Validation and Feature Selection [26] showed that the combination of cross-validation and variable selection significantly improved the robustness of predictions.
- A comparative study Integrated Deep Learning Model for Heart Disease Prediction [27] evaluated the performance of neural networks, support vector machine (SVM), and logistic regression, highlighting the trade-offs between accuracy and interpretability.

This work is aligned with our methodology and research objectives, which include:

1. Developing a predictive model for heart failure from a structured dataset of clinical characteristics.
2. Preprocessing and exploring the dataset to identify trends and optimize feature representation.
3. Applying feature engineering to transform raw data into meaningful variables.
4. Optimizing performance through cross-validation and hyperparameter tuning.

5. Training and evaluating the model using performance metrics such as accuracy, AUC, and the confusion matrix.
6. Ensuring interpretability and clinical utility, essential for deployment in decision-support tools.

2.1 Potential of big data analytics and AI in healthcare diagnoses

The potential of big data analytics and AI in healthcare diagnoses is significant, as evidenced by the literature.

Big data analytics enables the analysis of large volumes of healthcare data, supporting evidence-based decision-making and personalized medicine [3].

AI technologies, such as deep learning and machine learning [12], have shown promising results in the screening, diagnosis, and treatment of various diseases, indicating the potential for AI to significantly impact medical diagnosis and therapy [13].

Furthermore, the application of AI in the development of intelligent decision support systems has been highlighted, emphasizing the role of AI in generating decision alternatives to aid decision-makers [14].

The potential of big data analytics and AI in healthcare diagnoses is further supported by the transformative impact of AI and big data in healthcare decision support systems [15].

The synthesis of these references underscores the significant potential of big data analytics and AI in enhancing healthcare diagnoses through the development of intelligent diagnostic decision support systems.

3 PROPOSED RESEARCH METHOD

In this section, we aim to develop a predictive model for heart failure using a dataset containing various clinical features. The dataset used for this analysis is the "heart.csv" file provided by Kaggle [9], which includes information about patients' age, sex, blood pressure, cholesterol levels, and other relevant attributes. The project involves exploratory data analysis, feature engineering, model training, and evaluation.

3.1 Data preprocessing and exploratory data analysis

The implementation begins with the importation of the required libraries, followed by loading and exploring the dataset using PySpark [16].

Figure 1 displays its schema as loaded in PySpark. It lists all features, their data types, and nullability, which are essential for guiding the preprocessing and feature engineering steps.

```
root
 |-- Age: integer (nullable = true)
 |-- Sex: string (nullable = true)
 |-- ChestPainType: string (nullable = true)
 |-- RestingBP: integer (nullable = true)
 |-- Cholesterol: integer (nullable = true)
 |-- FastingBS: integer (nullable = true)
 |-- RestingECG: string (nullable = true)
 |-- MaxHR: integer (nullable = true)
 |-- ExerciseAngina: string (nullable = true)
 |-- Oldpeak: double (nullable = true)
 |-- ST_Slope: string (nullable = true)
 |-- HeartDisease: integer (nullable = true)
```

Fig. 1. Schema of the heart failure dataset

During the exploratory data analysis phase, we gained valuable insights into the dataset. Some key observations include:

- The dataset contains various features such as age, sex, blood pressure, cholesterol levels, and other clinical attributes.
- The target variable, "HeartDisease," indicates whether a patient has heart disease or not.
- The distribution of features such as age, blood pressure, cholesterol levels, and maximum heart rate (MaxHR) was visualized using histograms and box plots.
- Count plots were used to analyze the distribution of categorical variables such as sex, chest pain type, and fasting blood sugar.

The visualizations in Figure 2 provided us with a better understanding of the entire dataset and helped us identify potential patterns or relationships between the features and the target variable.

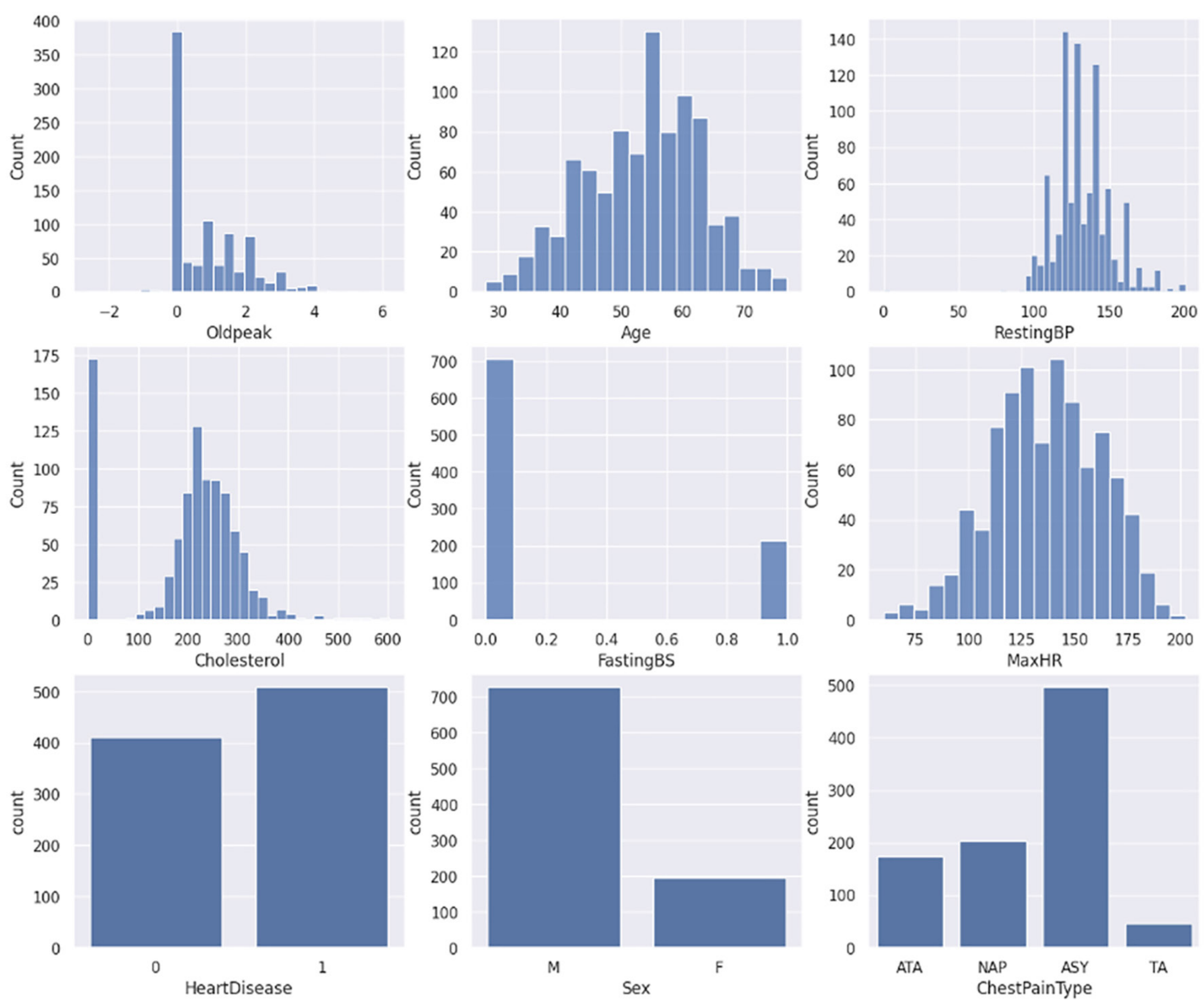

Fig. 2. *(Continued)*

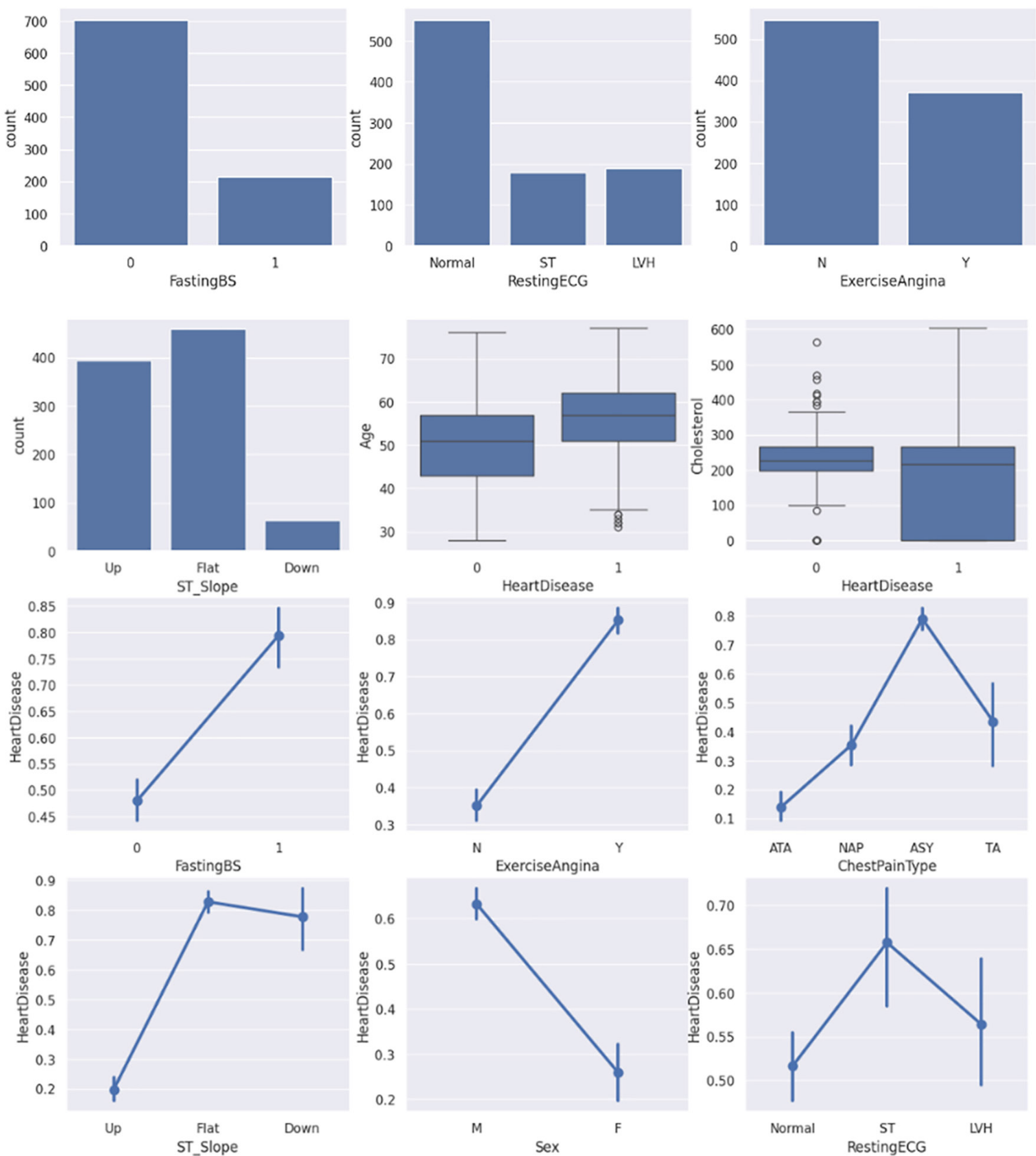

Fig. 2. Data visualization for exploratory data analysis

3.2 Data cleaning and feature engineering

The dataset was thoroughly checked for missing values, and no null values were found. Therefore, no imputation or removal of records was necessary.

Feature engineering techniques were applied to transform the data and create new features. Categorical variables such as sex, chest pain type, and fasting blood sugar were converted into binary indicators using custom transformations. This process ensured that the categorical variables could be effectively utilized in the predictive model.

Figure 3 is a pair plot matrix that visualizes the relationships between different continuous variables present in our dataset.

Columns and rows: Each box represents the relationship between two variables.

Colored scatterplots:

- Orange (1): Represents individuals with heart disease.
- Blue (0): Represents those without heart disease.

Main diagonal: Shows individual variable distributions via density graphs (KDE).

Relationship between MaxHR and heart disease: It appears that a drop in maximum heart rate (MaxHR) correlates with the presence of heart disease (orange dots concentrated on lower values).

Old peak: there is a slight separation between the two classes (orange and blue dots), indicating that higher values of "Oldpeak" (a measure of ST-segment depression) may be associated with heart disease.

Cholesterol and Resting BP: no clear distinction between the two classes for these variables. There was a wide dispersion, with high cholesterol values present in both groups (sick and not sick).

Fasting BS (fasting blood glucose): the variable is binary (0 or 1). The distribution shows that the number of patients with high blood sugar levels is slightly more frequent in people with heart disease.

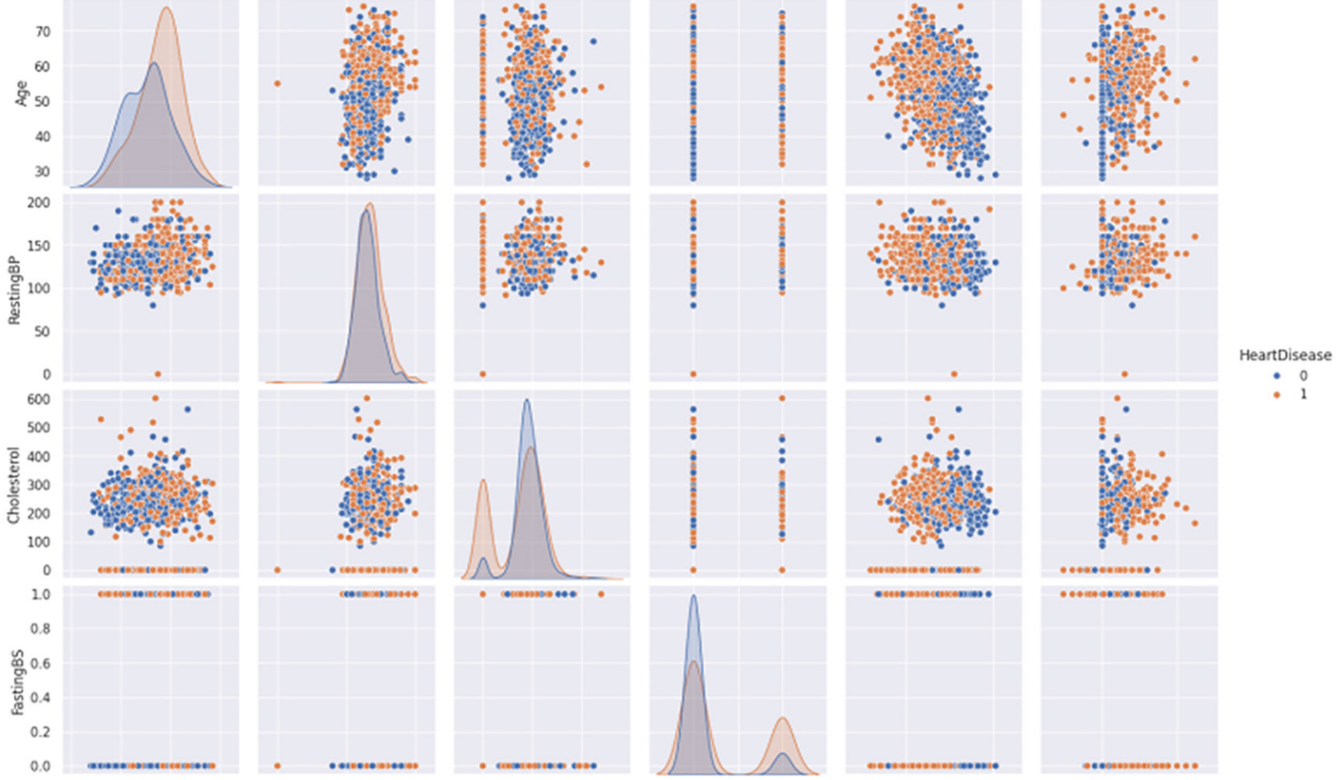

Fig. 3. *(Continued)*

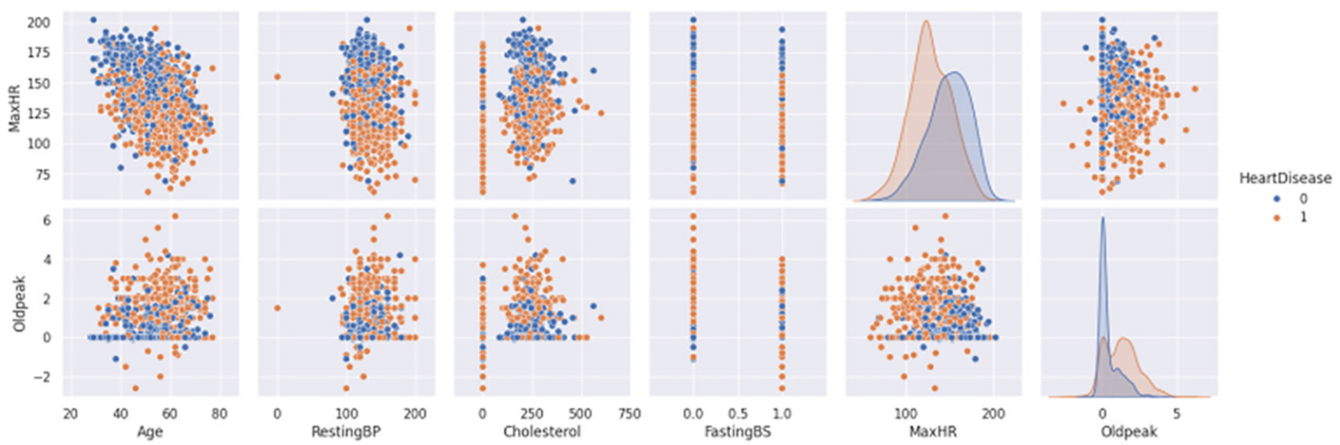

Fig. 3. Data distribution with and without heart disease

Figure 4 is a correlation matrix showing the linear relationships between the variables in the dataset.

This heatmap displays the pairwise Pearson correlation coefficients between numerical features in the dataset. Values close to +1 or –1 indicate strong linear relationships. The color gradient helps to visually identify correlations that may be relevant for feature selection or model interpretation.

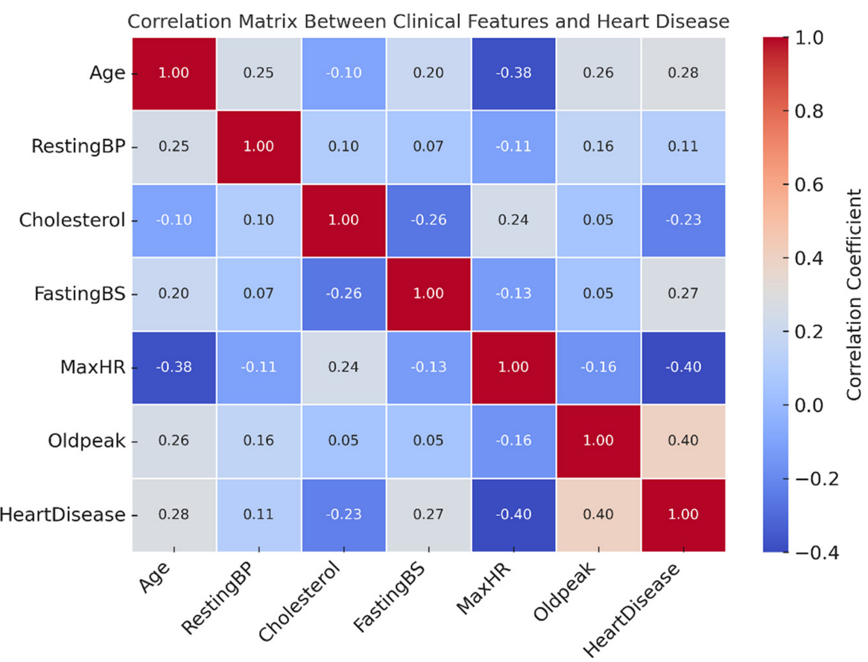

Fig. 4. Correlation matrix between clinical variables and heart disease

As illustrated in Figure 4, the feature *Oldpeak* shows a moderate positive correlation with heart disease, while *MaxHR* is negatively correlated. These insights guide the choice of relevant predictors for the model.

Figure 5 illustrates the transformation of categorical and numerical features into a format compatible with machine learning models. The dataset has been encoded and structured to optimize predictive performance in the classification of heart disease.

```
+----------------------------------------------------------+-----+
|features                                                  |label|
+----------------------------------------------------------+-----+
|(12,[0,1,2,4,7,9],[40.0,140.0,289.0,172.0,1.0,1.0])       |0    |
|(12,[0,1,2,4,5,10],[49.0,160.0,180.0,156.0,1.0,1.0])      |1    |
|[37.0,130.0,283.0,0.0,98.0,0.0,0.0,1.0,0.0,1.0,0.0,1.0]   |0    |
|[48.0,138.0,214.0,0.0,108.0,1.5,1.0,0.0,1.0,0.0,1.0,0.0]|1    |
|(12,[0,1,2,4,9],[54.0,150.0,195.0,122.0,1.0])             |0    |
+----------------------------------------------------------+-----+
only showing top 5 rows
```

Fig. 5. Visualization of feature transformation applied for heart failure prediction

To meet the requirements of the machine learning algorithms, the dataset shown in Figure 6 was prepared. This is a crucial step in guaranteeing optimal model performance and minimizing the biases associated with non-uniform data distributions.

	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	label	yes_exercise_angina	ata_chest_pain_type	asy_chest_pain_type	up_st_slope	f_sex	st_resting_ecg
0	40	140	289	0	172	0.0	0	0	1	0	1	0	0
1	49	160	180	0	156	1.0	1	0	0	0	0	1	0
2	37	130	283	0	98	0.0	0	0	1	0	1	0	1
3	48	138	214	0	108	1.5	1	1	0	1	0	1	0
4	54	150	195	0	122	0.0	0	0	0	0	1	0	0

Fig. 6. Final feature-engineered dataset for model training

3.3 Train-test split

The transformed dataset is split into training and testing sets using a 70:30 ratio. The count of records in each set is displayed. The analysis of the dataset reveals that there are 359 occurrences of the value '1' in the data, representing approximately 53.90% of the total instances. This reflects a relatively balanced dataset, which is suitable for binary classification tasks. In particular, the training set was used to perform a 5-fold cross-validation on the best-performing algorithm, enabling a more reliable estimation of its generalization capacity.

The final evaluation on the test set was conducted only once, after model selection, ensuring a fair and realistic assessment of the model's predictive power.

As part of the training of our heart disease prediction model, the distribution of data after the train-test split is shown in Figure 7.

```
1 train.select("label").count()
```

666

```
1 test.select("label").count()
```

252

Fig. 7. Data split for model training and evaluation

3.4 Model training and evaluation

Multi-algorithm comparison. To identify the most effective predictive model for heart failure detection, a series of machine learning algorithms were individually evaluated using the weighted F1-score—a balanced metric that accounts for both precision and recall, weighted by class frequency.

The evaluated models include:

- Decision tree
- XGBoost
- Random forest
- SVM
- Logistic regression

Each model was trained and evaluated separately while maintaining strict consistency in data handling. Specifically, all algorithms were applied to the same preprocessed dataset, using identical exploratory analysis steps, categorical feature transformations, vector assembly, and pipeline configurations. This uniformity ensures that the observed performance differences are solely attributable to the characteristics of the algorithms themselves, rather than any variation in data preparation.

The comparative graph shown in Figure 8 highlights the following results:

- Logistic regression achieved the highest weighted F1-score (0.87).
- SVM and random forest followed closely with a score of 0.86.
- XGBoost scored 0.85, while the decision tree model had the lowest score (0.81).

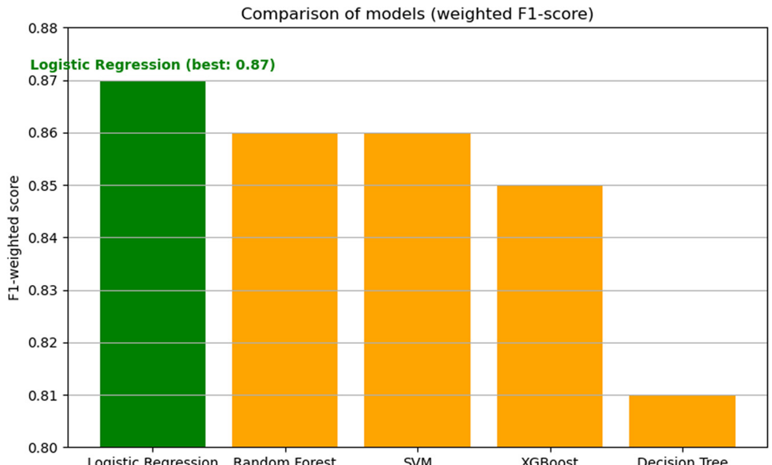

Fig. 8. Comparison of machine learning models

Based on these results, logistic regression was selected for the next phase of in-depth evaluation via cross-validation. This choice was guided not only by its strong performance but also by its robustness and interpretability—key qualities in clinical decision-making contexts.

Cross-validation of the selected model. To assess the robustness of the selected logistic regression model, a 5-fold cross-validation was conducted on the training set. A grid search over the regularization parameter (regParam) values ranging from 0.0 to 1.0 was performed to identify the optimal level of regularization.

The cross-validation results, shown in Figure 9, reveal that the best mean F1-score (0.844) was achieved at regParam = 0.0, indicating that no regularization yielded the optimal performance for these data. The model trained with this setting reached an even higher F1-score of 0.87 when evaluated on the separate test set, demonstrating excellent generalization capability.

This analysis confirms that regularization was not necessary for improving the model performance and highlights the model's stability and reliability in predicting heart failure outcomes.

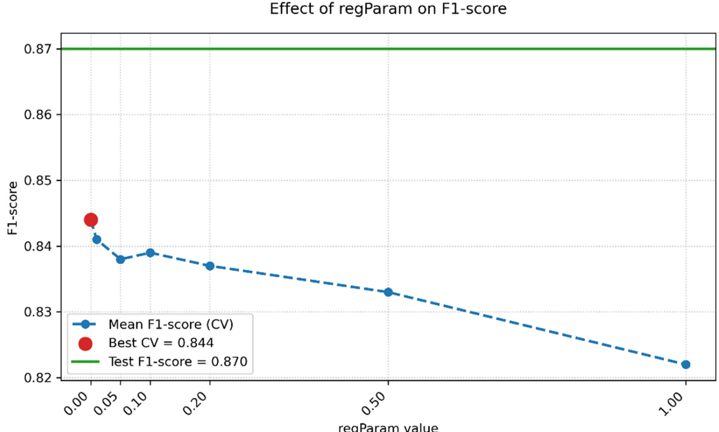

Fig. 9. Cross-validation vs test F1-scores for different regParam values

Final training of the best model. Based on the cross-validation results, the final logistic regression model was retrained on the entire training dataset using the optimal configuration (regParam = 0.0). This full training maximizes the use of available data and prepares the model for final evaluation on the hold-out test set.

To further illustrate the model's discriminative performance and generalization ability, the ROC curves obtained from both full training and 5-fold cross-validation are presented in Figure 10.

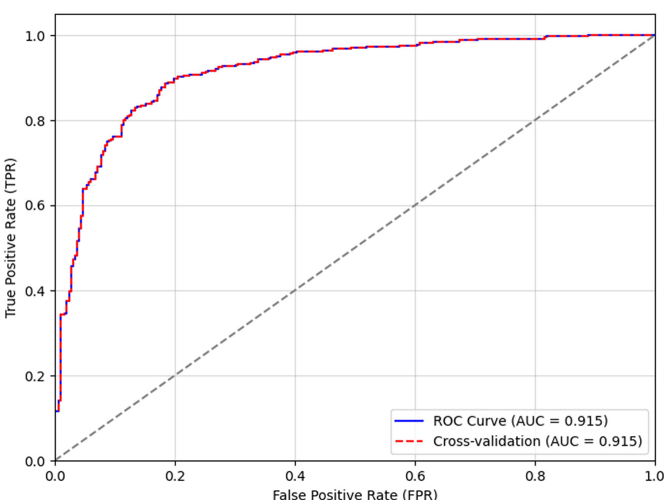

Fig. 10. Roc curve of the final logistic regression model

The consistently high AUC values (0.915) observed in both settings confirm the model's strong discriminative power and stability, validating its ability to generalize effectively to unseen data.

4 RESULTS AND DISCUSSION

4.1 Evaluation of the intelligent diagnostic decision support system

To evaluate the performance of the intelligent diagnostic decision support system [17], a classification report was generated in Table 1 using PySpark.

The performance achieved by the final model is not only quantitatively strong but also aligned with the clinical requirements of a decision support system. Logistic regression proved to be the most effective algorithm among those tested, while offering essential interpretability in a sensitive medical context.

Unlike more complex models such as XGBoost or random forest, logistic regression enables direct understanding of the weights assigned to each variable, which facilitates adoption by healthcare professionals. This transparency is crucial to foster user trust in the system's predictions.

The results confirm the model's excellent predictive capacity:

- **Final F1-score: 0.87**, indicating strong consistency between precision and recall.
- A detailed **classification report** further demonstrates robust performance across both classes in terms of precision and recall.

Table 1. Classification report

Target Class	Precision	Recall	F1-Score	Support
0.0	0.87	0.83	0.85	112
1.0	0.87	0.90	0.88	140
Weighted avg	0.87	0.87	0.87	252

A confusion matrix used to evaluate the performance of a binary classification model is illustrated in Figure 11.

- Vertical axis (True label): the true classes (real labels).
- Horizontal axis (Predicted label): classes predicted by the model.

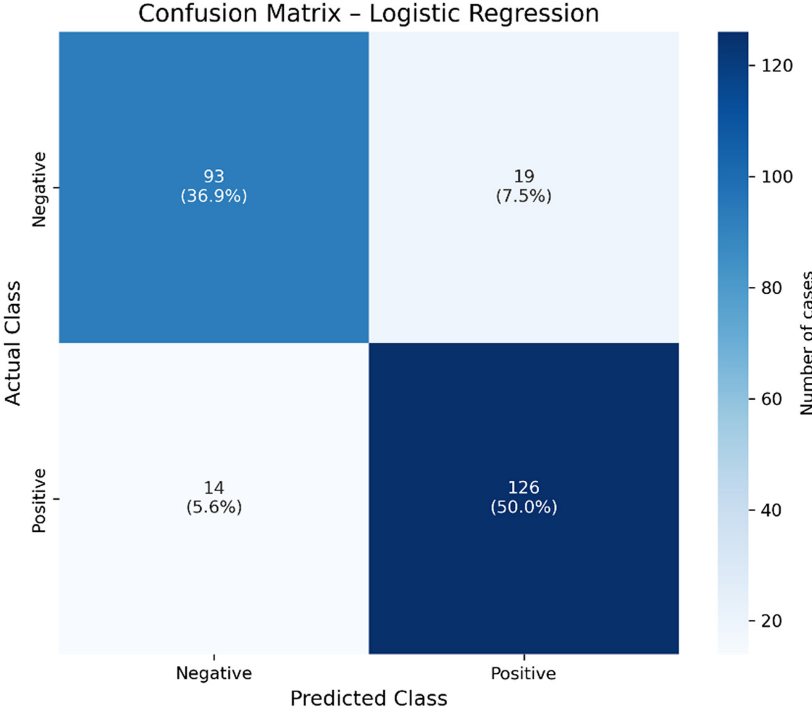

Fig. 11. Confusion matrix for the logistic regression model

The confusion matrix shows that the final logistic regression model correctly identified 93 out of 112 negative cases (specificity ≈ 83%) and 126 out of 140 positive cases (sensitivity ≈ 90%). The false negative rate (5.6%) and false positive rate (7.5%) remain low, indicating a well-balanced predictive performance.

This confirms the model's ability to **minimize both critical diagnostic errors**: missing actual positive cases (false negatives) and misclassifying healthy individuals (false positives). Such performance is particularly relevant in clinical settings, where both types of errors carry significant consequences.

To better understand the key predictive factors in heart failure detection, the relative importance of each feature was derived from the coefficients of the final logistic regression model. The following Figure 12 displays the normalized contribution of each variable within the model.

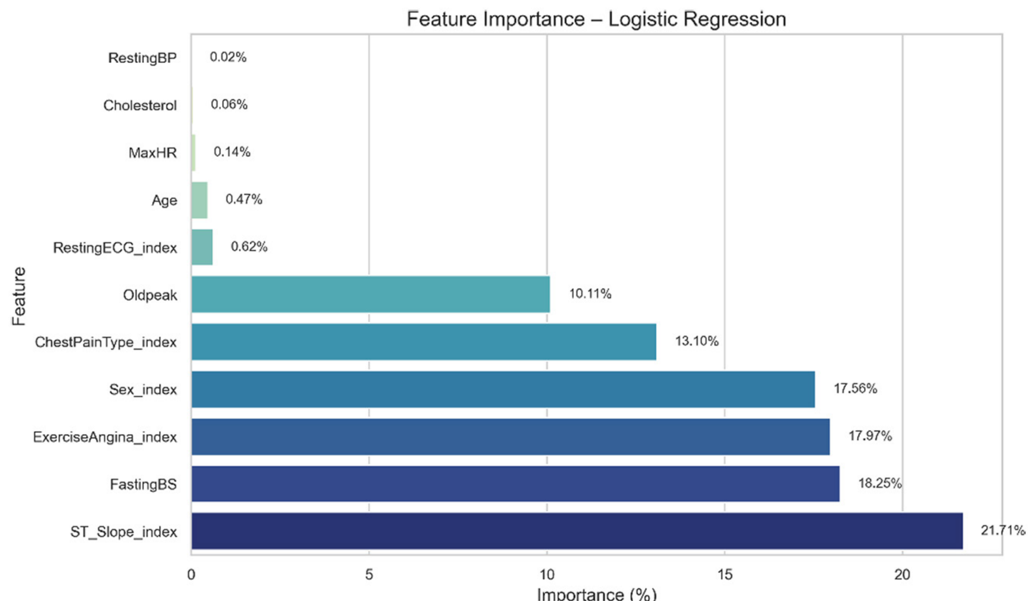

Fig. 12. Feature importance for heart failure diagnosis

The figure highlights the predominant role of variables such as *ST_Slope_index*, *FastingBS*, and *ExerciseAngina_index*, which together account for a significant portion of the total importance. These results confirm the clinical relevance of these features in heart failure prediction and illustrate the model's ability to effectively prioritize key variables.

4.2 System architecture for clinical decision support

The conceptual architecture presented shows how the IDDSS can be deployed in a real clinical setting.

Figure 13 highlights the end-to-end data flow, from user input to the final prediction.

In the envisioned configuration, a healthcare professional enters a patient's clinical data through a graphical interface. These data are then converted into a format compatible with a machine-learning model (logistic regression). The model produces a prediction, which is returned to the user in a clear, interpretable form.

This architecture serves as a reference framework demonstrating the technical feasibility of embedding a machine-learning model within an interactive medical interface. Although the system has not yet been rolled out in a hospital environment, the design outlines the pathway for eventual implementation.

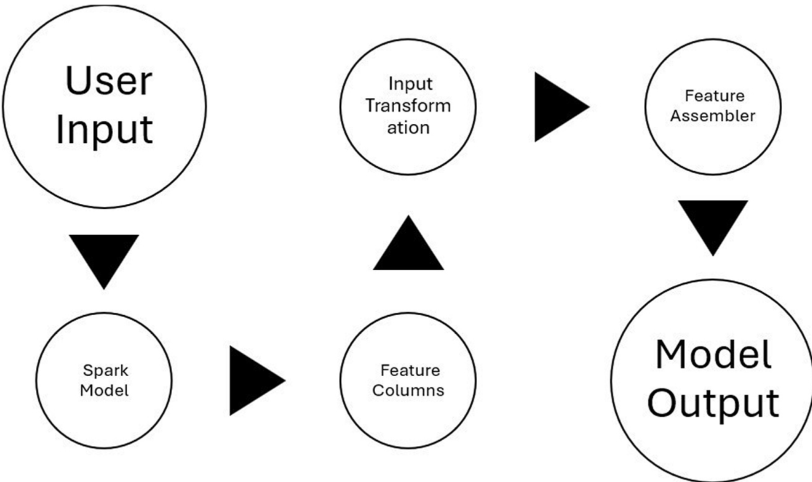

Fig. 13. Conceptual web architecture of the IDDSS

Technical workflow depicted in Figure 13:

1. The user enters values for various heart-health-related features.
2. The input values are sent to the Spark application.
3. Spark loads the pre-trained logistic regression classification model.
4. Feature columns are defined.
5. The user data are converted into a Spark DataFrame.
6. Categorical string columns in the DataFrame are indexed with **StringIndexer**.
7. Feature columns are assembled into a single feature vector with **VectorAssembler**.
8. The transformed data are passed to the model for prediction.
9. The prediction result is displayed to the user.

4.3 Using the model to predict streaming data from IoT devices

In this segment, we'll explore ways we can use the trained logistic regression model to predict streaming data from IoT devices, such as smartwatches or fitness bands. Streaming data refers to a continuous flow of data that is generated in real-time by IoT devices.

To predict streaming data, we need to establish a connection between the IoT devices and the system where the logistic regression model is deployed. This can be achieved by implementing a data ingestion pipeline that receives the streaming data from the devices and processes it in real-time.

The pipeline should preprocess the incoming streaming data to match the format expected by the logistic regression model. This may involve data cleaning, feature extraction, or scaling, depending on the specific requirements of the model.

Once the streaming data is preprocessed, it can be passed through the logistic regression model to obtain predictions. The model should be able to handle the continuous flow of data and make predictions in real time.

The data flow illustrated in Figure 14 shows how AI-based systems and distributed technologies can transform healthcare by enabling proactive monitoring and fast action. The integration of IoT, Kafka, and Spark ensures fluid, real-time data processing for critical decisions.

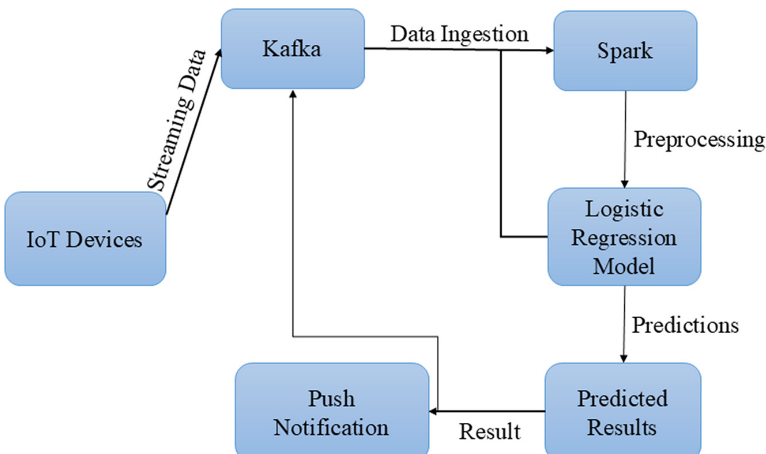

Fig. 14. Data flow in the heart failure prediction system

The predicted results can then be utilized for various purposes, such as monitoring the health status of individuals wearing the IoT devices, detecting anomalies or abnormalities, or triggering alerts or notifications based on specific conditions. It is important to ensure that the system has the necessary infrastructure and resources to handle the incoming streaming data and make predictions in real-time. This may involve scaling up the computational resources or implementing distributed processing techniques to handle the high volume and velocity of the streaming data.

By leveraging the trained logistic regression model in this manner, we can enable real-time prediction and monitoring using streaming data from IoT devices, opening up possibilities for applications in healthcare, fitness tracking, and other domains [18].

4.4 Deployment and practical integration

Integrating the IDDSS into on-premises hospital environments, or directly in the cloud, relies on a modular architecture illustrated by our processing pipeline. The workflow begins with a data-entry (or automated ingestion) layer that collects patient data, structured according to standards such as HL7 or FHIR, from hospital information systems (HIS). These data then undergo a transformation phase (cleaning, normalization, and typing), followed by categorical-variable encoding with a StringIndexer, represented by the Input Transformation block (see Figure 15).

The transformed data are aggregated into a single feature vector by a **Feature Assembler**, ensuring compatibility with the Spark-based prediction model (the **Spark Model** block). The model produces a prediction, which is either displayed in a user interface or forwarded to a third-party system as an FHIR resource (*Observation* or *DiagnosticReport*). The pipeline concludes with the model output, which can feed a dashboard, a clinical API, or a decision-support data store.

For production deployment, this architecture can be containerized into microservices (via **Docker**) and orchestrated with **Kubernetes** on cloud platforms

such as AWS, Azure, or GCP. Integration with HIS is achieved through secure interoperability gateways (FHIR Gateway, Kafka Connect HL7), enabling data exchanges that comply with privacy regulations (**GDPR**, **HIPAA**) and ensuring seamless synchronization with real-time clinical data flows.

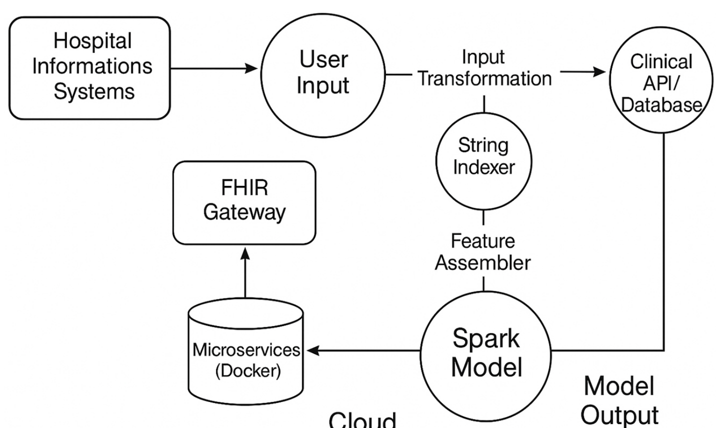

Fig. 15. Integration architecture of the IDDSS pipeline within a hospital environment

4.5 Challenges and opportunities associated with the integration of big data analytics and AI in healthcare diagnoses

A) Challenges

- **Data privacy and security:** The use of big data analytics and AI in healthcare raises concerns about patient data privacy and security. Healthcare organizations must ensure that patient data is protected and compliant with data privacy regulations.
- **Data quality:** Healthcare data is often complex and heterogeneous, leading to challenges in data quality and consistency. Ensuring the accuracy and reliability of healthcare data is crucial for the success of big data analytics and AI applications.
- **Interoperability:** Healthcare data is often stored in disparate systems and formats, making interoperability a significant challenge [19]. Integrating and harmonizing data from different sources is essential for effective big data analytics and AI in healthcare [20].
- **Ethical considerations:** The use of AI in healthcare diagnoses raises ethical considerations, such as bias in algorithms, transparency of decision-making, and accountability for AI-driven diagnoses.
- **Regulatory compliance:** Healthcare organizations must comply with regulatory requirements when implementing big data analytics and AI in healthcare. This includes adherence to data protection laws, patient consent, and ethical guidelines.
- **Resource constraints:** Implementing big data analytics and AI in healthcare requires significant resources, including infrastructure, expertise, and investment. Healthcare organizations, particularly in resource-constrained settings, may face challenges in adopting these technologies.
- **Clinical adoption:** Healthcare professionals may face challenges in adopting and trusting AI-driven diagnoses [21]. Ensuring the acceptance and adoption

of AI technologies by clinicians is essential for successful integration in healthcare [22].

B) Opportunities

- **Enhanced diagnoses:** Big data analytics and AI have the potential to enhance healthcare diagnoses by leveraging large volumes of data to identify patterns, trends, and insights that may not be apparent through traditional methods.
- **Personalized medicine:** AI-driven diagnoses enable personalized and precision medicine [23], tailoring treatments and interventions to individual patient characteristics and needs.
- **Early detection:** AI technologies can facilitate early detection of diseases and conditions, leading to timely interventions and improved patient outcomes.
- **Clinical decision support:** AI-driven decision support systems provide valuable insights and recommendations to healthcare professionals [24], aiding in clinical decision-making and patient management.
- **Public health surveillance:** Big data analytics and AI enable real-time monitoring of public health trends, disease outbreaks, and epidemiological patterns, supporting public health surveillance and response.
- **Research and innovation:** The integration of big data analytics and AI in healthcare diagnoses fosters research and innovation, driving advancements in medical science, technology, and patient care.
- **Healthcare access:** AI technologies can improve access to healthcare services [25], particularly in underserved and remote areas, by enabling telemedicine, remote monitoring, and virtual consultations.

C) Ethical and legal considerations

The dataset used in this study is the publicly available "Heart Failure Prediction" dataset from Kaggle, which is anonymized and intended for academic use. Nevertheless, in a real-world clinical deployment of the IDDSS, strict compliance with data protection regulations such as GDPR and HIPAA would be essential. This would involve mechanisms for obtaining informed consent, applying pseudonymization techniques, tracking data access, and securing the infrastructure using encryption protocols and robust authentication. These measures are critical to ensuring the ethical and secure handling of sensitive health data in practical applications.

5 CONCLUSION

This study highlights the power of artificial intelligence and big data techniques in developing an intelligent decision support system for heart failure diagnosis. Among the evaluated algorithms, logistic regression emerged as the most effective model, offering an excellent balance between predictive performance and interpretability, which is essential in clinical settings.

By combining cross-validation, hyperparameter tuning, and feature importance analysis, the final model demonstrates strong robustness and generalization capacity, as confirmed by its performance on independent test data. The analysis of key features further validates the clinical relevance of the system, in line with current medical knowledge.

This work opens up concrete opportunities, including extending the system to other chronic diseases, integrating multi-source data, and deploying the model in connected hospital environments for real-time decision support. Continued research and development in this direction have the potential to reshape healthcare

practices and support the rise of a more proactive, personalized, and data-driven precision medicine.

6 REFERENCES

[1] A. Oussous, A. Ez-Zahout, S. Ziti, and A. Oussous, "An overview of the most efficient methods for predicting healthcare disorders," *AIP Conference Proceedings*, vol. 2814, no. 1, p. 020003, 2023. https://doi.org/10.1063/5.0150025

[2] M. Huang, H. Han, H. Wang, L. Li, Y. Zhang, and U. A. Bhatti, "A clinical decision support framework for heterogeneous data sources," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 6, pp. 1824–1833, 2018. https://doi.org/10.1109/JBHI.2018.2846626

[3] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare," *BioMed. Res. Int.*, vol. 2015, no. 1, pp. 1–16, 2015. https://doi.org/10.1155/2015/370194

[4] Y. Wang, L. Kung, and T. A. Byrd, " Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technological Forecasting and Social Change*, vol. 126, pp. 3–13, 2018. https://doi.org/10.1016/j.techfore.2015.12.019

[5] P. S. Ramapraba, M. Radhika, S. Sumathi, J. Karthik, and N. Senthamilarasi, "An efficient healthcare system by cloud computing and clustering-based hybrid machine learning algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 3, pp. 1698–1707, 2024. https://doi.org/10.11591/ijeecs.v34.i3.pp1698-1707

[6] Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," in *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, Hefei, China, 2016, pp. 1383–1388. https://doi.org/10.1109/ICIEA.2016.7603801

[7] Q. Long, X. Ye, and Q. Zhao, "Artificial intelligence and automation in valvular heart diseases," *Cardiology Journal*, vol. 27, no. 4, pp. 404–420, 2020. https://doi.org/10.5603/CJ.a2020.0087

[8] N. R. Kolukula, P. N. Pothineni, V. M. K. Chinta, V. G. Boppana, R. P. Kalapala, and S. Duvvi, "Predictive analytics of heart disease presence with feature importance based on machine learning algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, pp. 1070–1077, 2023. https://doi.org/10.11591/ijeecs.v32.i2.pp1070-1077

[9] Fedesoriano, "Heart failure prediction dataset," Kaggle, 2021. Accessed: Feb. 17, 2024. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

[10] A. Ez-zahout, H. Gueddah, A. Nasry, R. Madani, and F. Omary, "A hybrid big data movies recommendation model based k-nearest neighbors and matrix factorization," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 434–441, 2022. https://doi.org/10.11591/ijeecs.v26.i1.pp434-441

[11] N. Narisetty, A. Kalidindi, M. V. Bujaranpally, N. Arigela, and V. V. Ch, "Ameliorating heart diseases prediction using machine learning technique for optimal solution," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 16, pp. 156–165, 2023. https://doi.org/10.3991/ijoe.v19i16.42071

[12] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data*, vol. 9, 2022. https://doi.org/10.1186/s40537-022-00592-5

[13] A. Mitsala, C. Tsalikidis, M. Pitiakoudis, C. Simopoulos, and A. K. Tsaroucha, "Artificial intelligence in colorectal cancer screening, diagnosis and treatment. A new era," *Current Oncology*, vol. 28, no. 3, pp. 1581–1607, 2021. https://doi.org/10.3390/curroncol28030149

[14] A. Najjar, B. Amro, and M. Macedo, "An intelligent decision support system for recruitment: Resumes screening and applicants ranking," *Informatica*, vol. 45, no. 4, pp. 617–623, 2021. https://doi.org/10.31449/inf.v45i4.3356

[15] A. Tiwari, "Transforming paradigm of artificial intelligence for scalable clinical healthcare," *Advanced Materials Letters*, vol. 12, no. 7, pp. 1–3, 2021. https://doi.org/10.5185/amlett.2021.071642
[16] N. Nelamali, "PySpark add a new column to dataframe," Spark By {Examples}, 2024. https://sparkbyexamples.com/pyspark/pyspark-add-new-column-to-dataframe/
[17] R. Madani, A. Idrissi, and A. Ez-Zahout, "A new context-based factorization machines for context-aware recommender systems," in *Modern Artificial Intelligence and Data Science: Tools, Techniques and Systems*, vol. 1102, A. Idrissi, Éd., Springer, Cham, 2023, pp. 15–23. https://doi.org/10.1007/978-3-031-33309-5_2
[18] C. Iwendi, S. Khan, J. H. Anajemba, A. K. Bashir, and F. Noor, "Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model," *IEEE Access*, vol. 8, pp. 28462–28474, 2020. https://doi.org/10.1109/ACCESS.2020.2968537
[19] Z. Yang, "Design of intelligent decision support system based on artificial intelligence," in *3rd International Conference on Internet of Things and Smart City (IoTSC 2023)*, 2023, pp. 374–378. https://doi.org/10.1117/12.2683893
[20] D. Muhunzi, L. Kitambala, and H. Mashauri, "Big data analytics in the healthcare sector: Opportunities and challenges in developing countries. A literature review," *Research Square*, 2023. https://doi.org/10.21203/rs.3.rs-2869049/v1
[21] P. Cordero, M. Enciso, D. López, and A. Mora, "A conversational recommender system for diagnosis using fuzzy rules," *Expert Systems with Applications*, vol. 154, p. 113449, 2020. https://doi.org/10.1016/j.eswa.2020.113449
[22] M. Schmude *et al.*, "Investigating the potential for clinical decision support in sub-Saharan Africa with AFYA," *JMIR Research Protocols*, vol. 11, no. 6, pp. 1–17, 2022. https://doi.org/10.2196/34298
[23] W. Abbaoui, S. Retal, N. Kharmoum, and S. Ziti, "Artificial intelligence at the service of precision medicine," in *International Conference on Advanced Intelligent Systems for Sustainable Development, AI2SD 2022,* in Lecture Notes in Networks and Systems, J. Kacprzyk, M. Ezziyyani, and V. E. Balas, Eds., vol. 713, Springer, Cham, 2023, pp. 91–103. https://doi.org/10.1007/978-3-031-35248-5_9
[24] A. Macharla and G. Pradeepini, "A novel framework for the diagnosis of Parkinson's disease using transfer learning with RESNET50 and SVM classifier," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, pp. 877–886, 2023. https://doi.org/10.11591/ijeecs.v32.i2.pp877-886
[25] M. Suleiman, H. Demirhan, L. Boyd, F. Girosi, and V. Aksakalli, "A clinical coding recommender system," *Knowledge-Based Systems*, vol. 210, p. 106455, 2020. https://doi.org/10.1016/j.knosys.2020.106455
[26] I. E. Guabassi, Z. Bousalem, R. Marah, and A. Haj, "Artificial neural networks with K-fold cross-validation and feature selection for early heart disease prediction," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 14, pp. 102–115, 2024. https://doi.org/10.3991/ijoe.v20i14.51479
[27] S. A. H. Fazlur and S. K. Thillaigovindan, "Integrated deep learning model for heart disease prediction using variant medical data sets," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 18, no. 9, pp. 178–191, 2022. https://doi.org/10.3991/ijoe.v18i09.30801

7 AUTHORS

Ghlana Amara holds a professional degree in IT Development from ISCAE, Mauritania (2019), and a master's degree in Data Engineering and Software Development from the Faculty of Sciences, Mohammed V University (UM5),

Morocco (2021). She is currently pursuing a Ph.D. in Artificial Intelligence at the Department of Computer Science, UM5. With over five years of experience in developing digital solutions, her expertise spans data analysis, IT security and ethics, and the optimization of recommendation systems. Her research interests focus on intelligent systems and AI-based applications in healthcare (E-mail: ghlana_amara@um5.ac.ma).

Abderrahmane Ez-Zahout currently holds the position of an Assistant Professor at the Department of Computer Science within the Faculty of Sciences at Mohammed V University. He specializes in computer science, digital systems, big data, and computer vision. Lately, his focus has shifted towards intelligent systems. He has contributed as a reviewer for numerous scholarly journals and actively participates in NGOs, student associations, and the management of a non-profit foundation (E-mail: abderrahmane.ezzahout@um5.ac.ma).

Soumia Ziti Dr. Ziti is a Full Professor and researcher at the Faculty of Sciences of Mohammed V University in Rabat since 2007. She obtained her PhD in computer science specializing in graph theory from the University of Orleans in France, along with a diploma in advanced studies in fundamental computer science. She also holds a Baccalaureate in Mathematical Sciences and completed her undergraduate studies in mathematics and physics, specializing in mathematics, at Hassan II University in Morocco. Furthermore, she earned a master's degree in science and technology in computer science from the same institution. Her research interests encompass a wide range of topics including graph theory, information systems, artificial intelligence, data science, software development, database modelling, big data, cryptography, and numerical methods and simulations. Pr. Ziti has contributed extensively to these fields with over than eighty publications in esteemed international journals and conferences. Additionally, she plays a pivotal role in coordinating, participating or assessing in various educational and socio-economic or research projects (E-mail: s.ziti@um5r.ac.ma).