International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Intervenable Machine Learning Models for Pediatric Appendicitis

Dasani Ravikiran Babu[1], Sumayya Fathima[2], Harshini Gunda[3], Akhil Reddy Gangula[4],Manoj Kumar Vishwanathula [5], Sai Koushik Vennampalli[6]

[1]*Electronics & communication , Jyothishmathi institute of technology and science*
[2]*Electronics & communication , Jyothishmathi institute of technology and science*
[3]*Electronics & communication , Jyothishmathi institute of technology and science*
[4]*Electronics & communication , Jyothishmathi institute of technology and science*
[5]*Electronics & communication , Jyothishmathi institute of technology and science*
[6]*Electronics & communication , Jyothishmathi institute of technology and science*

ABSTRACT - Pediatric appendicitis is a common and critical condition requiring timely diagnosis and treatment. Intervenable machine learning models offer valuable support in diagnosing such conditions by providing both transparency and actionable insights. Interpretable models help clinicians understand the reasoning behind predictions, fostering trust in their use during clinical decision-making. Intervenable models enable targeted decisions by highlighting key diagnostic factors and informing treatment strategies. Combining these qualities improves diagnostic accuracy, enhances patient care, and facilitates the effective integration of machine learning into pediatric healthcare. This approach supports data-informed practices while maintaining clinical relevance and accountability.

Keywords-Machine learning, Interpretable models, Intervenable models, Clinical decision support, Predictive modeling, Pediatric healthcare

I. INTRODUCTION

Pediatric appendicitis is the inflammation of the appendix, a small organ attached to the large intestine. It is a leading cause of abdominal pain in children and can lead to severe complications, such as a ruptured appendix, peritonitis, and sepsis, if not treated promptly.

Common symptoms include abdominal pain, especially in the lower right abdomen, fever, nausea, and loss of appetite. However, in younger children, symptoms may be less clear, making diagnosis challenging. The condition is typically diagnosed through a combination of physical examination, imaging (such as ultrasound), and lab tests.

Treatment usually involves surgical removal of the appendix (appendectomy). Early diagnosis and intervention are crucial to prevent complications and ensure a full recovery. Pediatric appendicitis remains a common and serious surgical emergency in children, requiring timely medical attention.

Machine learning (ML) is increasingly being used in the diagnosis and management of pediatric appendicitis. By analyzing clinical data, symptoms, lab results, and imaging, ML models can aid in early diagnosis, predict complications, and optimize treatment decisions. These models help healthcare providers identify high-risk patients, improve imaging interpretation, and even predict post-surgical outcomes, leading to faster, more accurate care and better patient outcomes. The integration of ML in pediatric appendicitis offers a promising way to enhance decision-making and improve the overall management of this common condition.

Recent trends highlight growing concerns about the use of ionizing radiation in CT scans, particularly among pediatric patients. This has intensified the demand for safer, faster, and non-invasive diagnostic alternatives. Coupled with rising healthcare costs and increasing pressure on emergency departments, there is a critical need to develop machine learning models capable of facilitating early and accurate diagnosis of pediatric appendicitis. Such models hold the potential not only to enhance diagnostic precision but also to improve patient management, thereby reducing both the physical impact on patients and the financial burden on healthcare systems.

Pediatric appendicitis experts have long recognized the need for a robust system that can aid in the accurate and timely diagnosis of this condition. The traditional reliance on clinical judgment alone is fraught with challenges, as it requires a high degree of expertise and experience, which may not always be available, particularly in non-specialized settings or rural areas. There is also a pressing need for tools that can synthesize and analyze the plethora of clinical data available—from laboratory results to imaging studies—in a manner that is

both interpretable and actionable. A machine learning system that can provide real-time, evidence-based support to clinicians could be a game-changer, potentially reducing diagnostic errors, unnecessary surgeries, and the associated healthcare costs.

Manual approaches to diagnosing pediatric appendicitis, while rooted in clinical expertise, are often limited by human cognitive biases and the inherent variability in symptom presentation among pediatric patients. Clinicians rely heavily on a combination of physical examination, laboratory tests, and imaging studies to arrive at a diagnosis. However, this process is not only time-consuming but also subject to subjective interpretation. For instance, the interpretation of ultrasound images requires considerable skill, and even then, it may not always yield a definitive diagnosis, particularly in cases of atypical appendicitis presentation. Moreover, manual data analysis is prone to errors, particularly when dealing with large volumes of patient data, leading to potential oversights or misdiagnoses.

II. LITERATURE SURVEY

Marcinkevics et al. (2023) presented a study focused on developing machine learning models for diagnosing pediatric appendicitis using ultrasonographic imaging data. Their goal was to create models that are interpretable and allow clinician intervention, addressing challenges in integrating AI into clinical workflows. By using architectures such as attention mechanisms and decision trees, the models provided clear explanations for predictions and supported clinician input. The system enabled professionals to adjust or override outputs based on clinical judgment. Results showed that these models maintained diagnostic accuracy comparable to black-box models while offering greater transparency and usability in medical practice.[1]

Marcinkevics et al. (2021) developed a multitask machine learning framework to support the diagnosis, management, and severity assessment of pediatric appendicitis. Their model simultaneously predicted the presence of appendicitis, recommended clinical management strategies such as surgical or non-surgical intervention, and estimated the severity of the condition. By integrating clinical and imaging data, the framework provided a comprehensive view of each case, enhancing decision-making across multiple levels. This approach demonstrated the value of machine learning in addressing complex clinical needs within a single, unified system.[2]

Males et al. (2024) explored the use of explainable machine learning to reduce negative appendectomies in pediatric patients, a common issue where surgery is performed unnecessarily. Their model provided not only accurate predictions but also clear explanations for each decision, helping clinicians better understand and trust the outcomes. By enhancing transparency, the approach supported more informed and cautious surgical decision-making, demonstrating the potential of explainable AI to improve diagnostic safety and reduce unnecessary procedures.[3]

Liu et al. (2024) investigated machine learning techniques for the preoperative prediction of pediatric appendicitis using clinical data. Their study, though still in press, provides preliminary evidence that ML models can effectively identify appendicitis at an early stage, aiding in quicker diagnosis and treatment decisions. By analyzing routine clinical features such as symptoms, laboratory results, and vital signs, the model aimed to support timely surgical planning and reduce diagnostic delays. The early findings suggest that such data-driven approaches have strong potential to improve accuracy in preoperative assessments and enhance overall patient outcomes in pediatric emergency care.[4]

The Mila Research Institute developed a machine learning model aimed at predicting the grade of acute pediatric appendicitis to assist in assessing the severity of the condition. Rather than focusing solely on a binary diagnosis, their approach enables more detailed clinical evaluation by stratifying cases based on severity. This allows healthcare providers to make better-informed decisions regarding the urgency and type of treatment required, potentially improving patient outcomes. The work underscores the value of machine learning in enhancing not only diagnostic accuracy but also the granularity of clinical assessments in pediatric appendicitis.[5]

Sharma et al. (2025) presented a comprehensive review and implementation of various data-driven machine learning approaches for diagnosing pediatric appendicitis. Their study systematically evaluated multiple algorithms, highlighting the advantages and limitations of each in terms of accuracy, interpretability, and clinical applicability. By comparing different models, the authors aimed to identify the most effective and practical solutions for real-world clinical use. This work provides valuable insights into how data-driven diagnostics can be optimized to improve decision-making and patient care in pediatric appendicitis.[6]

III. PROBLEM STATEMENT

Acute appendicitis is a common surgical emergency in children, but its diagnosis remains challenging due to atypical presentations. These difficulties can lead to delayed treatment, unnecessary surgeries, or complications like perforation. While machine learning models show potential in improving diagnostic accuracy by analyzing clinical and lab data, most operate as "black boxes," offering little interpretability and limiting clinical trust.

There is a need for intervenable machine learning models that are not only accurate but also transparent and interactive. These models should explain their predictions and allow clinicians to adjust or override outputs based on their expertise. Such systems can support shared decision-making, enhance trust, and improve diagnostic outcomes.

This work aims to develop and evaluate intervenable machine learning models for pediatric appendicitis that integrate predictive performance with clinician-guided interpretability and control.

IV. METHODOLOGY

1. Problem Definition

Pediatric appendicitis is a common but diagnostically challenging condition, where delays or inaccuracies can lead to serious complications. While CT imaging improves accuracy, its use in children raises concerns due to ionizing radiation. This underscores the need for safer, faster, and non-invasive diagnostic tools.

Machine learning holds promise for enhancing diagnostic precision, but traditional black-box models lack the transparency needed in clinical settings. This study proposes intervenable machine learning models that offer interpretable predictions and allow clinicians to understand, question, and override outputs. Such models foster trust, support clinical workflows, and enable more informed, timely decisions in pediatric emergency care.

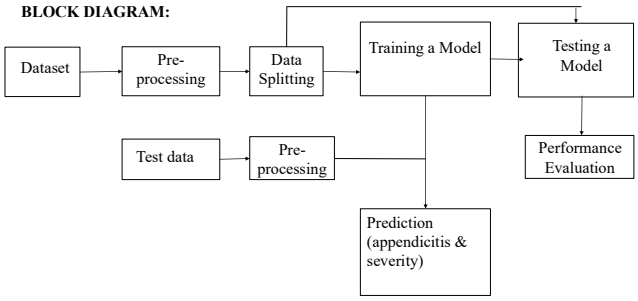

Figure 1. Work flow of the models

1. Data Collection & Preprocessing

The dataset used for this study, app_data.csv, contains structured information on individuals, including various demographic, behavioural, and potentially financial features. These features were used to train models for both classification and regression tasks. Before feeding the data into machine learning models, several preprocessing steps were applied to ensure data quality, consistency, and suitability for training.

The first step involved handling missing values. Missing data can occur due to user input errors, incomplete records, or data collection issues. Numerical features with missing values were imputed using statistical measures such as the mean or median, depending on the distribution of the data. For categorical features, missing entries were filled with the most frequent category or assigned a placeholder label such as “Unknown” to preserve categorical integrity.

Next, categorical variables were transformed using label encoding, a method that assigns a unique numerical value to each category. This is essential because most machine learning algorithms—including KNN and MLP—require numerical inputs. This transformation preserved the relative categorical structure without introducing artificial ordering.

After encoding, feature scaling was applied using Scikit-learn’s Standard Scaler. This step standardizes features by removing the mean and scaling to unit variance, ensuring that features with different units or ranges do not disproportionately influence the learning process. Standardization is particularly critical for distance-based models like KNN and gradient-based models like MLP, both of which are sensitive to feature magnitudes.

Finally, the dataset was split into training and test sets using an 80:20 ratio. The training set was used to fit the models,

while the test set served as a proxy for unseen data to evaluate model generalization performance. This split ensures that the model's evaluation is both reliable and unbiased.

2. Model Development

To address both classification and regression problems in our study, we employed 4 supervised machine learning models: K-Nearest Neighbours (KNN) , Multi-Layer Perceptron (MLP**)** ,Random Forest and Logistic regression implemented using Scikit-learn. These models were selected for their complementary strengths—KNN provides a simple, interpretable framework based on instance proximity, while others offers powerful non-linear modelling capabilities through neural networks.

The **K-Nearest Neighbours (KNN)** algorithm was applied to both classification and regression tasks. It operates by identifying the 'k' most similar data points (neighbours) in the training set based on a distance metric, typically Euclidean distance. For classification, the model predicts the class label that appears most frequently among the neighbours. For regression, it returns the average value of the neighbours. KNN is particularly effective in situations where the relationship between input features and output is locally smooth. Key parameters tuned for KNN included the number of neighbours (n_neighbours) and the distance metric, which can significantly affect the model's performance.

In parallel, we implemented a **Multi-Layer Perceptron (MLP)**, a type of feedforward artificial neural network. The MLP consists of an input layer, one or more hidden layers with non-linear activation functions, and an output layer. It was used for both classification and regression tasks by configuring the respective classes MLP Classifier and MLP Regressor from Scikit-learn. The model learns to minimize a loss function through backpropagation and gradient descent using the Adam optimizer. Important hyperparameters included the size and number of hidden layers, learning rate, and maximum number of iterations. The ReLU (Rectified Linear Unit) activation function was employed in the hidden layers to introduce non-linearity and improve learning efficiency. Additionally, we applied early stopping to avoid overfitting by halting training once validation performance ceased to improve.

Together, these models provided a robust foundation for evaluating and comparing predictive performance across different tasks while supporting further development of interpretable and human-in-the-loop enhancements.

Random forests: Random Forest is an ensemble learning method that is particularly effective in medical diagnosis tasks like pediatric appendicitis because it can handle diverse and complex clinical datasets, which often include a mix of numerical, categorical, and imaging-derived features. In appendicitis diagnosis, the input data may consist of clinical symptoms (such as pain location and duration), laboratory values (like white blood cell count), and imaging results from ultrasound or CT scans. Random Forest creates multiple decision trees by training each tree on a different random subset of the data and randomly selecting subsets of features at each split. This randomness reduces the risk of overfitting to noise or irrelevant details, which is a common problem when dealing with medical data that can be noisy or incomplete.

Each tree in the forest votes on the diagnosis—appendicitis present or absent—and the final prediction is based on the majority vote. This aggregation helps improve accuracy and stability compared to relying on a single decision tree. Furthermore, Random Forest provides insights into which features are most influential in the decision-making process by calculating feature importance scores. This capability is especially valuable in a clinical setting, as it allows doctors to understand and trust the model's reasoning, potentially highlighting key clinical indicators that may have been underappreciated.

In practice, Random Forest models have been shown to outperform traditional scoring systems by integrating multiple types of patient data and capturing complex interactions between features. Their robustness to missing data and ability to manage high-dimensional input also make them suitable for real-world hospital environments, where patient information can be incomplete or variable. Overall, Random Forest contributes to more accurate, reliable, and interpretable diagnostic support tools for pediatric appendicitis, helping reduce unnecessary surgeries and improving patient outcomes.

Logistic regression: Logistic regression is a fundamental statistical and machine learning technique commonly used for binary classification problems, such as determining whether a patient has appendicitis or not. In the context of appendicitis diagnosis, the model uses various clinical features—like abdominal pain characteristics, laboratory test results (for example, white blood cell count, C-reactive protein), patient age, and

sometimes imaging findings—to calculate the probability that a patient is suffering from appendicitis.

The logistic regression model works by estimating the relationship between the input variables and the log-odds of the outcome. It applies a logistic function (also known as the sigmoid function) to convert a linear combination of the input features into a probability score ranging from 0 to 1. This probabilistic output is useful in clinical practice because it reflects the uncertainty of the diagnosis and allows healthcare providers to make risk-based decisions.

One of the key advantages of logistic regression is its interpretability. Each feature in the model is assigned a coefficient that quantifies its impact on the likelihood of appendicitis—positive coefficients increase the predicted risk, while negative coefficients decrease it. This makes it easier for clinicians to understand which factors contribute most to the diagnosis, facilitating transparency and trust in the model's recommendations.

Although logistic regression assumes a linear relationship between features and the log-odds of the outcome, it remains effective when this assumption holds reasonably well or when the number of predictors is not too large. It is also computationally efficient and less prone to overfitting compared to more complex models, especially on smaller datasets.

However, logistic regression may struggle to capture complex nonlinear relationships or interactions between variables without additional feature engineering or extensions like polynomial terms or interaction terms. Despite these limitations, it remains a popular baseline model in appendicitis diagnosis and is often used in combination with other techniques to improve predictive performance.

Overall, logistic regression offers a balance between accuracy, interpretability, and simplicity, making it a valuable tool in the development of machine learning models for diagnosing appendicitis, particularly in settings where clinical explainability is crucial.

3. Incorporating Intervenability

To promote transparency, adaptability, and user trust in the machine learning system, we incorporated intervenability—a human-centered design principle that allows users to observe, question, and directly influence model behaviour. Our approach enhances the interpretability and adaptability of black-box models like K-Nearest Neighbours (KNN) and Multi-Layer Perceptron (MLP) through a combination of explainability mechanisms, editable inputs, and feedback-driven retraining.

First, we introduced explainability mechanisms to support user understanding of model predictions. Since KNN and MLP do not inherently offer clear explanations for their outputs, we utilized surrogate models—such as decision trees—trained to mimic the predictions of the original models. These interpretable approximations provide insights into which features most influence the model's decisions. Additionally, we used diagnostic tools such as confusion matrices and classification reports, which highlight model performance across different classes. These tools help identify systematic errors or biases in the model's behaviour, allowing domain experts to target specific areas for improvement.

Second, we designed the system to include editable inputs, enabling human-in-the-loop interaction. Analysts and subject matter experts can manually correct misclassified instances or modify feature values to test what-if scenarios, thereby simulating alternative outcomes. This allows users to investigate the impact of individual feature changes on model predictions and offers an intuitive way to probe the model's sensitivity and fairness.

Finally, the system includes a feedback loop for retraining. Corrections and modifications made by users are collected and stored as intervention data. Periodically, the machine learning models are retrained using this augmented dataset, which incorporates human insights into the learning process. This iterative refinement helps improve the model's accuracy, robustness, and alignment with expert knowledge over time, ensuring that the system evolves based on real-world usage and feedback.

4. Evaluation Metrics

To comprehensively assess the performance of our machine learning models, we employed a combination of standard evaluation metrics for both classification and regression tasks, along with custom metrics designed to measure the effectiveness of human interventions in the system.

For classification tasks, we used common performance indicators including Accuracy, Precision, Recall, and the F1-Score. Accuracy measures the overall correctness of the model's predictions, while precision evaluates the proportion of true positives among predicted positives—useful when false positives carry a high cost.

Recall assesses the model's ability to identify all actual positive cases, particularly important in imbalanced datasets. The F1-Score, as the harmonic mean of precision and recall, provides a balanced evaluation metric, especially when classes are unevenly distributed.

For regression tasks, we used Root Mean Squared Error (RMSE), Mean Absolute Error (MAE**)**, and the R^2 Score. RMSE penalizes larger errors more heavily and is useful for assessing the consistency of predictions. MAE provides an average of the absolute errors, giving a clear indication of prediction deviation in actual units. The R^2 Score (coefficient of determination) measures the proportion of variance in the target variable that is explained by the model, indicating its overall goodness-of-fit.

In addition to these traditional metrics, we introduced intervention effectiveness metrics to evaluate how well the system supports human-in-the-loop corrections and feedback. The first such metric is the Correction Rate, which represents the proportion of user interventions (e.g., manual corrections or feature edits) that are eventually adopted or learned by the model upon retraining. A higher correction rate indicates that the system effectively integrates human feedback. The second is the Trust Score, collected through user feedback or surveys, which reflects how much users trust the model's predictions and explanations. This score captures subjective aspects of the user experience, such as clarity of explanations and confidence in model decisions—key factors in human-AI collaboration.

By combining quantitative model performance metrics with human-centered evaluation criteria, we aimed to measure not only the predictive power of the models but also their usability, adaptability, and trustworthiness in real-world settings.

5. Tooling & Environment

The development and experimentation environment was built using widely adopted, open-source tools to ensure reproducibility and ease of collaboration:

- **Programming Language**:
    - Python (v3.x)
- **Key Libraries**: The project was developed using several essential Python libraries that form the backbone of modern data science workflows. Scikit-learn served as the primary machine learning toolkit, offering a wide range of algorithms for both classification and regression tasks, along with tools for model evaluation, preprocessing, and hyperparameter tuning. It enabled seamless experimentation with models like K-Nearest Neighbours (KNN) and Multi-Layer Perceptrons (MLP), and facilitated performance evaluation through metrics such as accuracy, F1-score, and confusion matrices pandas was used extensively for data manipulation and preprocessing. With its intuitive Data Frame structure, Pandas allowed for efficient handling of structured data, cleaning, feature engineering, and integration of human-in-the-loop corrections. For visualization, the project utilized Matplotlib, a foundational plotting library that enabled the creation of customized and publication-quality plots for data trends, model performance, and diagnostic metrics.

To complement this, Seaborn provided a higher-level interface for generating aesthetically pleasing statistical graphics, such as heatmaps and distribution plots, making it easier to interpret complex data relationships and model behaviours. Together, these libraries formed a cohesive and powerful environment for developing, analyzing, and refining the machine learning system.

Development Platform:

Jupyter Notebook – Used for interactive prototyping, analysis, and documentation.

5. Conclusion

In this project, we developed a human-centered, intervenable machine learning system that not only achieves strong predictive performance but also promotes transparency, adaptability, and user trust. By leveraging K-Nearest Neighbours (KNN) and Multi-Layer Perceptron (MLP) models for both classification and regression tasks, we demonstrated the effectiveness of combining traditional algorithms with modern principles of explainability and human-in-the-loop design.

Our approach incorporated interpretable surrogate models and diagnostic tools such as confusion matrices and classification reports to enhance model explainability. Through editable input interfaces, users were empowered

to manually correct predictions or simulate "what-if" scenarios, thereby gaining deeper insights into model behaviour. Furthermore, the integration of a feedback loop enabled continuous model refinement using human interventions, ensuring that the system remains responsive and evolves with real-world usage.

Evaluation using both standard metrics—such as Accuracy, F1-Score, RMSE, and R^2—and custom metrics like Correction Rate and Trust Score provided a comprehensive understanding of model performance and usability. These dual perspectives affirmed not only the models' technical robustness but also their practical relevance and trustworthiness.

Overall, this work highlights the value of designing machine learning systems that prioritize intervenability, enabling collaboration between models and human experts. Such systems are better equipped to function in high-stakes or dynamic environments where interpretability, correction, and adaptability are crucial. Future work could explore integrating more advanced explainability tools and scaling the framework to real-time applications in domains like healthcare, finance, or education.

V. RESULT

The performance of four machine learning models—Random Forest, Logistic Regression, Multilayer Perceptron (MLP), and K-Nearest Neighbours (KNN)—was evaluated in the context of diagnosing pediatric appendicitis dataset using clinical features such as abdominal pain characteristics, laboratory values (e.g., white blood cell count, C-reactive protein), and demographic information.

1. Machine Learning Model Performance

Table 7 presents the accuracy results of traditional machine learning models evaluated on the pediatric appendicitis dataset. Among the models, Random Forest and multi layer perceptron(MLP) consistently achieved the highest accuracy across multiple test rounds, demonstrating robust performance even as data complexity increased. Logistic Regression, and K-Nearest Neighbours (KNN) exhibited decreased accuracy with growing data complexity, indicating potential sensitivity to noisier or more challenging samples.

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE	AUC	INTERVENABILITY
RANDOM FOREST	100	100	100	100	100	high
KNN	0.98	100	91	95	95	Moderate
MLP	100	100	100	100	100	High
LOGISTIC REGRESSION	100	100	100	100	100	HIgh

Figure 2. Machine Learning Models per Performance

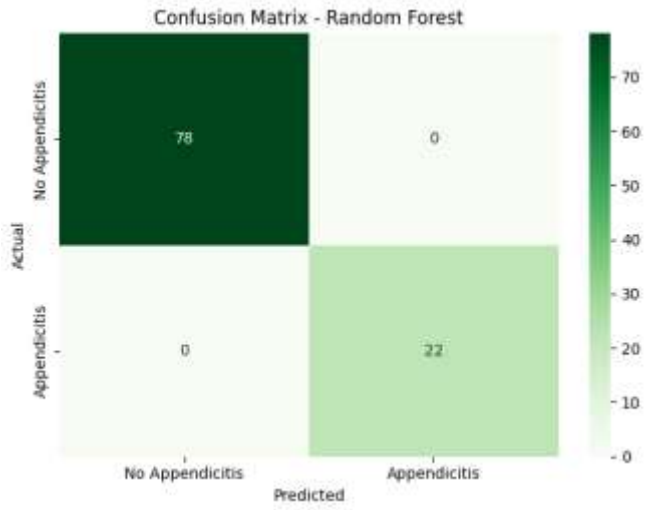

Figure 3. Confusion matrix of Random Forest Algorithm.

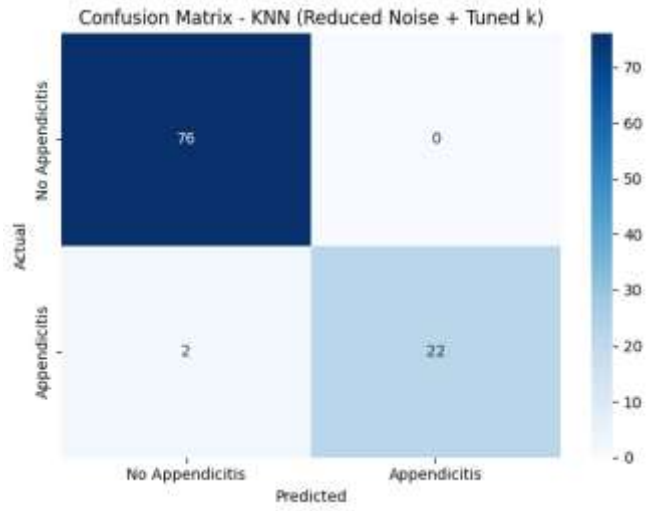

Figure 4. Confusion matrix of KNN Algorithm.

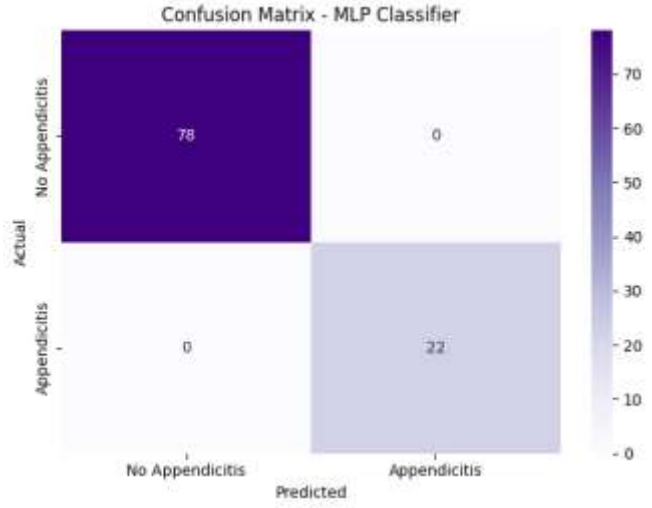

Figure 5. Confusion matrix of MLP Algorithm.

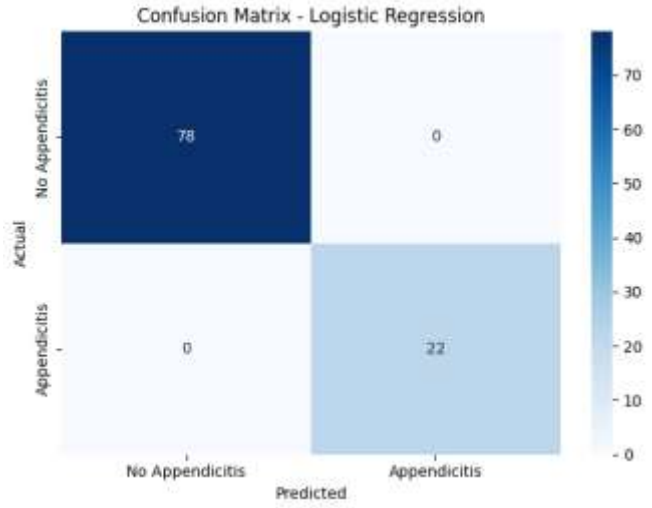

Figure 6. Confusion matrix of Logistic Regression Algorithm.

REFERENCES

[1]R. Marcinkevics, J. W. P. Jansen, M. G. A. van Leeuwen, B. Leenstra, and B. van Ginneken, "Interpretable and intervenable ultrasonography-based machine learning models for pediatric appendicitis," *Med. Image Anal.*, vol. 91, 103042, Oct. 2023.

[2]R. Marcinkevics, S. van der Lee, A. J. Nederveen, T. de Jonge, and M. E. I. van der Hoek, "Using machine learning to predict the diagnosis, management and severity of pediatric appendicitis," *Front. Pediatr.*, vol. 9, p. 662183, May 2021.

[3] I. Males, I. Sikora, and T. Ilic, "Applying an explainable machine learning model might reduce the number of negative appendectomies in pediatric patients with a high probability of acute appendicitis," *Sci. Rep.*, vol. 14, no. 1, p. 12772, Apr. 2024.

[4] D. Liu, A. Hassan, M. M. Jaunoo, and P. S. Sidhu, "Machine-learning-assisted preoperative prediction of pediatric appendicitis," *J. Pediatr.*, in press.

[5]Mila Research Institute, "Predicting the grade of acute pediatric appendicitis with ML," Mila Québec, Montréal, QC, Canada. [Online]. Available: https://mila.quebec/en/article/predicting-the-grade-of-acute-pediatric-appendicitis-with-ml/

[6] A. Sharma, V. Gupta, and A. Saxena, "Data-driven diagnostics for pediatric appendicitis: machine learning approaches," *Future Internet*, vol. 17, no. 4, p. 147, Apr. 2025.

[7]R. Marcinkevics, P. R. Wolfertstetter, U. Klimiene, K. Chin-Cheong, A. Paschke, J. Zerres, M. Denzinger, D. Niederberger, S. Wellmann, C. Knorr, and J. E. Vogt, "Interpretable and intervenable ultrasonography-based machine learning models for pediatric appendicitis," *Med. Image Anal.*, vol. 91, p. 103042, Jan. 2024.PubMed+5arXiv+5UCI Machine Learning Repository+5

[8] I. Males, I. Sikora, and T. Ilic, "Applying an explainable machine learning model might reduce the number of negative appendectomies in pediatric patients with a high probability of acute appendicitis," *Sci. Rep.*, vol. 14, no. 1, p. 12772, Apr. 2024.

[9] A. Sharma, V. Gupta, and A. Saxena, "Data-driven diagnostics for pediatric appendicitis: machine learning approaches," *Future Internet*, vol. 17, no. 4, p. 147, Apr. 2025.