AIntelia

Utilizing Machine Learning for Enhanced Diagnosis and Management of Pediatric Appendicitis: A Multilayer Neural Network Approach

İlyas Özer[1]

[1]*Bandırma Onyedi Eylül University, Dept. of Computer Engineering, Bandırma, Balıkesir, Turkey,*
iozer@bandirma.edu.tr, ORCID: 0000-0003-2112-5497

This study focuses on pediatric appendicitis, a leading cause of hospital admissions due to abdominal pain in children, characterized by a substantial risk of perforation, especially in younger patients. Traditional diagnostic methods, while effective, often lack specificity and are supplemented by varying laboratory and imaging techniques. This research introduces a novel application of machine learning (ML), specifically a multi-output neural network model, to address the complexities of diagnosing appendicitis, determining its severity, and guiding management strategies in pediatric cases. The model, with its unique architecture, has been trained and tested on a comprehensive dataset from Children's Hospital St. Hedwig in Regensburg, Germany, which includes a wide array of clinical data and ultrasound images. The results demonstrate remarkable accuracy in classifying management approaches, severity levels, and diagnosis, highlighting the model's potential in supporting clinical decision-making. While not a replacement for clinical judgment, this model serves as a promising tool in the ongoing efforts to improve pediatric appendicitis care, offering a glimpse into the future of AI-enhanced medical diagnostics and treatment planning.

Keywords: Pediatric Appendicitis, Multilayer Neural Network, Machine Learning

1. Introduction

This study delves into the critical area of pediatric appendicitis, a predominant cause of hospital admissions for abdominal pain in children [1]. Appendicitis represents a significant health concern, with a lifetime risk of 6 to 9% and a peak incidence in the 10 to 19-year age group [2]. A notable aspect of this condition is its higher perforation rates in preschool children compared to older children and adults, underscoring the importance of timely and accurate diagnosis and treatment.

Diagnosis of appendicitis traditionally relies heavily on clinical assessment, supplemented by laboratory data and imaging [3]. Serum biomarkers like the white blood cell (WBC) count and C-reactive protein (CRP) are useful indicators, as demonstrated by Acharya et al., who reported areas under the receiver operating characteristic (AUROC) curves of 0.75 and 0.80 for WBC count and CRP, respectively, in diagnosing acute appendicitis. However, the search for a specific biomarker for appendicitis in clinical practice remains elusive [4]. Abdominal ultrasonography, particularly of the appendix, is a standard, non-invasive imaging technique in children, though its effectiveness varies due to its operator-dependent nature. The Alvarado Score (AS) and Pediatric Appendicitis Score (PAS) are notable tools used in risk stratification, yet they are not routinely used in all clinical settings.

The management of acute appendicitis in children lacks consistent international guidelines. While minimally invasive appendectomy is the standard treatment, there is growing evidence supporting the effectiveness of conservative therapy with antibiotics [5,6]. Furthermore, reports of spontaneous resolution in uncomplicated cases suggest that an antibiotic-free approach could be viable in selected instances.

In this context, the application of machine learning (ML) in the medical field, particularly for early detection and monitoring, offers a transformative potential [7-11]. Supervised learning models, using large datasets, can discern complex statistical patterns predictive of specific outcomes. This study aims to leverage ML to achieve three crucial outcomes: diagnosing appendicitis, guiding management decisions (conservative vs. operative), and stratifying risk based on severity (such as gangrene and perforation). The objective is to develop and validate a pilot ML tool to assist physicians in diagnosing appendicitis at the initial presentation, assessing severity, and determining the appropriate management strategy.

This study represents a pioneering effort in using ML to simultaneously predict diagnosis, management (conservative vs. operative), and severity in pediatric appendicitis. While it's not intended to produce a finished clinical decision support system, it serves as a promising research prototype, potentially paving the way for more advanced, data-driven approaches in the diagnosis and treatment of appendicitis in children.

2. Preprocessing Pipeline

In the realm of data science and machine learning, preprocessing constitutes a critical phase where raw data is transformed into a format more suitable for modeling. This phase often involves several key steps, each tailored to enhance the overall quality and compatibility of the data with the chosen analytical model.

The first step in the preprocessing pipeline is the imputation of missing values. Missing data can significantly compromise the integrity of statistical analyses and predictive modeling. To address this issue, a common strategy employed is the substitution of missing values with the most frequent value observed within the same feature, known as the mode. This technique, referred to as "most frequent imputation," ensures the preservation of the dataset's underlying distribution while mitigating the introduction of bias.

Subsequent to the imputation step, categorical variables within the dataset are transformed using a process known as label encoding. Categorical data, which often presents in a non-numeric form (such as words or labels), needs to be converted into a numerical format to be effectively processed by machine learning algorithms.

Label encoding involves assigning a unique integer to each category within the feature. This process is distinct from one-hot encoding, where binary vectors are used to represent categories. Label encoding is particularly beneficial when the categorical variable denotes an inherent order or rank, but it's also a practical approach in scenarios where the computational simplicity and memory efficiency are of paramount concern.

Given the standard limitations of label encoders in handling multiple columns simultaneously, a custom transformer, MultiColumnLabelEncoder, is implemented. This transformer is designed to extend the functionality of the traditional label encoder, enabling it to process either multiple columns in a Pandas DataFrame or feature indices in a NumPy array. This versatility is crucial in ensuring that the transformer seamlessly integrates into a broader range of data processing pipelines, catering to diverse data formats encountered in practical scenarios.

The imputation and label encoding steps are integrated into a cohesive pipeline using Scikit-Learn's Pipeline tool. This approach encapsulates the sequential application of data transformation processes, promoting modularity and ease of replication. The pipeline architecture not only streamlines the preprocessing workflow but also enhances the reproducibility and scalability of the model development process.

In conclusion, the preprocessing pipeline, encompassing imputation and label encoding, plays a pivotal role in preparing raw data for sophisticated analyses in machine learning. The meticulous design of these steps ensures that the input data is rendered into a format that is both compatible with the algorithmic requirements and reflective of the underlying structure of the dataset. By addressing the challenges posed by missing values and categorical variables through informed strategies like most frequent imputation and label encoding, the preprocessing stage significantly contributes to the robustness and accuracy of predictive models. The integration of these processes into a unified pipeline further exemplifies the importance of systematic and efficient data preparation in the broader context of data-driven research and analysis.

3. Dataset

The dataset from the retrospective study at Children's Hospital St. Hedwig in Regensburg, Germany, presents a comprehensive collection of clinical data from a cohort of pediatric patients admitted with abdominal pain [12]. This extensive dataset is particularly remarkable for its inclusion of multiple abdominal B-mode ultrasound images for most patients, illustrating a diverse range of abdominal regions such as the right lower quadrant, appendix, intestines, lymph nodes, and reproductive organs. The variability in the number of ultrasound (US) views per patient, ranging from 1 to 15, reflects the thoroughness of the imaging approach and provides a rich visual insight into the patients' conditions.

In addition to these detailed ultrasonographic images, the dataset encompasses a wide array of clinical data. This includes results from laboratory tests, outcomes of physical examinations, and scores from clinical assessment tools like the Alvarado and pediatric appendicitis scores. These elements collectively offer a multifaceted view of each patient's health status, enhancing the depth of information available for analysis.

A notable aspect of this dataset is its classification of subjects with respect to three critical target variables: diagnosis (differentiating between appendicitis and no appendicitis), management (categorized as surgical versus conservative), and severity (classified as complicated, uncomplicated, or no appendicitis). This classification not only aids in understanding the immediate clinical decisions and outcomes but also serves as a valuable resource for broader medical research. Such categorization allows for the analysis of patterns and correlations between clinical, laboratory, and imaging data, and the eventual health outcomes, thereby offering insights into effective diagnostic and treatment strategies for pediatric abdominal pain.

The dataset's comprehensive nature, combining detailed imaging data with a broad spectrum of clinical information, makes it a valuable asset for medical research, particularly in improving the diagnosis and management of abdominal conditions in pediatric patients. It provides a unique opportunity to study the interplay between various diagnostic tools and to develop models that can enhance the accuracy and efficiency of medical diagnoses and interventions in pediatric care.

The dataset under examination offers a rich tapestry of patient demographics, clinical characteristics, and treatment outcomes, providing valuable insights into the patterns and profiles within a healthcare setting. A statistical summary of the data reveals that the average age of subjects is around 11.35 years, with a standard deviation of 3.53 years, indicating a moderate breadth in the age range. The Body Mass Index (BMI) averages approximately 18.91, reflecting a relatively normal weight range for the age group, albeit with a degree of variability (standard deviation of 4.39). When examining the length of hospital stays, the average duration is about 4.28 days. However, the standard deviation of 2.57 days suggests the presence of both brief and more extended stays.

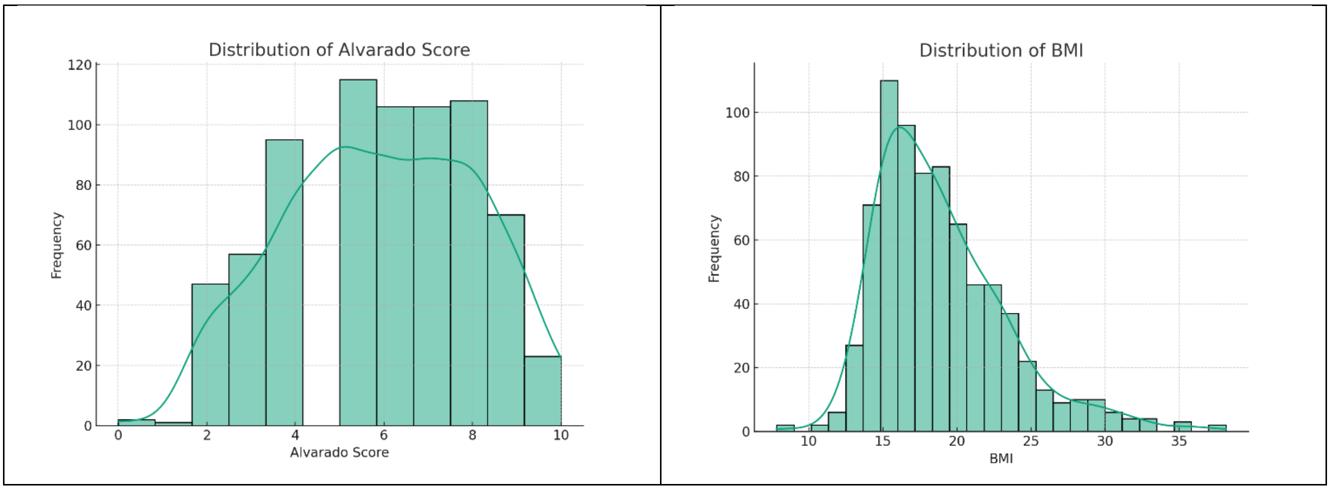
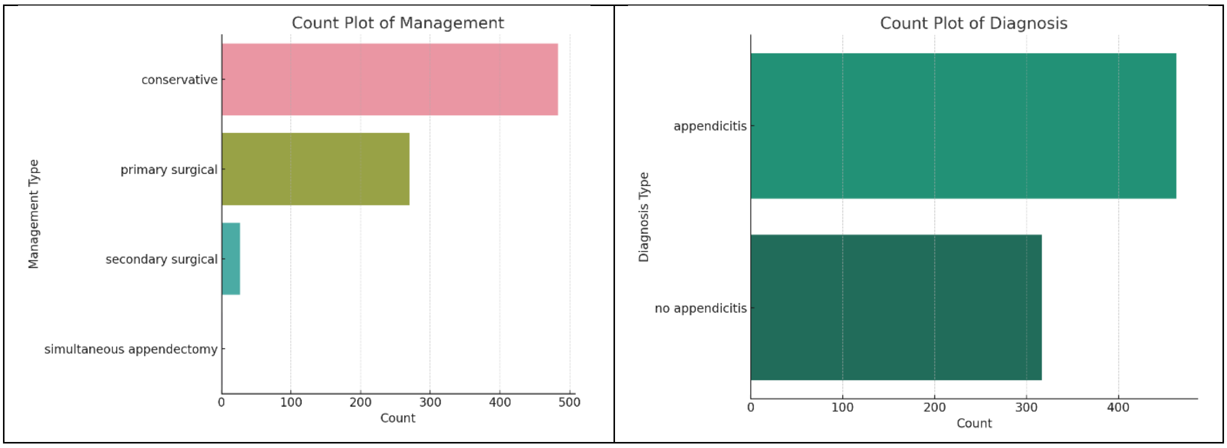


***Figure 1**. Statistical distribution of some features in the data set*

The Alvarado Score, instrumental in assessing the likelihood of appendicitis, averages near 5.92, with a range from 0 to 10. This range and a standard deviation of 2.16 point to a wide distribution in the severity of appendicitis symptoms among the patients. Another notable metric is the appendix diameter, which averages around 7.76 mm, but can extend up to 17 mm in some cases. Essential clinical parameters such as body temperature and blood parameters (like WBC Count and Neutrophil Percentage) exhibit expected averages, but also display notable variations, a common characteristic in clinical data reflecting the diverse health statuses of patients.

The visual analysis, comprising histograms and count plots, further elucidates these findings. Histograms for continuous variables such as age, BMI, length of stay, and Alvarado Score depict their respective distributions. For instance, the BMI histogram might indicate a normal distribution, whereas the length of stay histogram could show a right-skewed distribution, highlighting a predominance of shorter hospital stays. The count plots for categorical variables like management, severity, and diagnosis provide a clear visual representation of their frequency distribution within the dataset. These plots are instrumental in revealing how many cases fall under each category in management, the varying levels of severity, and the types of diagnoses made.

In conclusion, the combination of statistical and visual analysis of this dataset provides a comprehensive overview of patient demographics, clinical characteristics, and outcomes. Such in-depth analysis is vital for identifying underlying patterns, understanding patient profiles, and informing decision-making in both healthcare and research settings. This holistic view is crucial for advancing patient care and enhancing medical research.



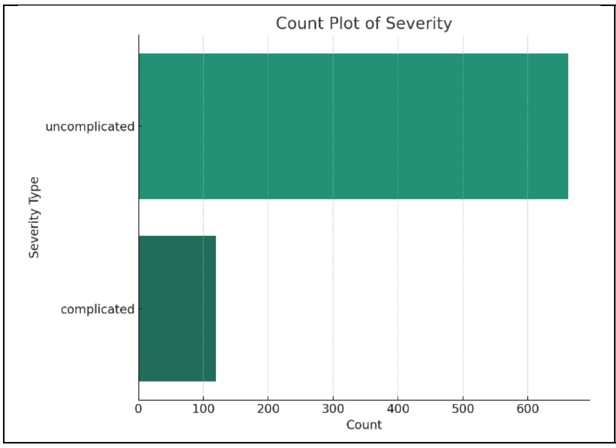


Figure 2. *Number of class information in the data set*

4. Multilayer Neural Network Model

In the realm of machine learning, the construction of a multi-output neural network model, as exemplified in the provided code, epitomizes a sophisticated approach to handling complex datasets, particularly those prevalent in clinical environments. This model is intricately designed to simultaneously address multiple classification tasks, a testament to the versatility and adaptability of neural networks in predictive analytics.

The architecture of the model is thoughtfully articulated to accommodate the distinct classification tasks—management, severity, and diagnosis—each representing a critical aspect of the clinical decision-making process. At the onset, the model commences with an input layer tailored to the dimensionality of the feature set derived from the transformed dataset. This layer serves as the foundational entry point for data processing.

Subsequent layers, consisting of two densely connected nodes with 64 and 32 neurons respectively, incorporate the ReLU activation function. This choice of activation function is pivotal in introducing non-linearity into the model, thereby empowering it to capture and learn more complex relationships within the data.

Diverging into task-specific pathways, the model features a trinity of output layers. The management prediction layer, equipped with three neurons, is aligned with the multiclass nature of the task, utilizing a softmax activation function to generate a probabilistic distribution across the potential classes. In contrast, the severity and diagnosis predictions are configured as binary classification tasks, each with a single-neuron layer employing a sigmoid activation function. This configuration is adept at producing a probability score indicative of class membership, suitable for binary classification scenarios.

The compilation phase of the model is meticulously calibrated with distinct loss functions for each output. The adoption of 'categorical_crossentropy' for the management task aligns with its multiclass classification nature, while 'binary_crossentropy' is judiciously applied to the binary classification tasks of severity and diagnosis. The inclusion of the Adam optimizer and accuracy as a metric in this phase reflects a strategic emphasis on balancing training efficiency and predictive effectiveness.

During the training regimen, the model undergoes fitting to the dataset, a process that involves the nuanced feeding of input features and corresponding target variables. This step is critical, as it involves the transformation of the target variables to formats congruent with the expectations of each output layer. Specifically, the management targets are one-hot encoded to complement the softmax output, while the severity and diagnosis targets are utilized in their existing form, befitting the sigmoid outputs. The model undergoes training over a course of 10 epochs with a batch size of 10, parameters that determine the iteration frequency over the dataset and the volume of data samples processed before each model update.

In conclusion, this neural network model exemplifies an advanced machine learning solution, adeptly engineered to navigate the intricacies of multi-output classification tasks within clinical datasets. Its architecture, training, and compilation strategies are a confluence of thoughtful design and methodical execution, underpinning the model's capability to unravel complex patterns in the data while distinctly addressing diverse classification tasks, integral to clinical decision-making.

5. Results & Discussion

In this study, the deployed neural network model demonstrates notable efficacy in classifying distinct clinical parameters within a medical dataset, encompassing management, severity, and diagnosis. The model's performance is quantified through accuracy and loss metrics, each offering a lens into its predictive capabilities and precision.

Regarding the classification of management strategies (surgical versus conservative), the model achieves a commendable accuracy of approximately 94.23%. This high degree of accuracy is indicative of the model's robustness in discerning between different management approaches. Complementing this, a loss value of around 0.142 signifies a low error rate in the model's predictions, reinforcing its effectiveness in this classification domain.

The model's proficiency is further exemplified in the severity classification task (complicated versus uncomplicated or no appendicitis), where it attains an accuracy of approximately 97.44%. The corresponding low loss value of 0.084 underscores the model's precision in categorizing the severity of the condition, a critical factor in clinical decision-making processes.

Most impressively, in the domain of diagnosis (appendicitis versus no appendicitis), the model exhibits an exemplary performance, achieving a perfect accuracy score of 100%. This exceptional accuracy, coupled with a remarkably low loss value of about 0.037, highlights the model's exceptional capability in accurately diagnosing the presence or absence of appendicitis.

These results collectively underscore the substantial potential of neural network models in complex classification tasks, especially within a medical context where precision and reliability are paramount. The consistently high accuracy and low loss across all categories suggest that the model has effectively captured the underlying patterns and nuances in the dataset, enabling it to make dependable and accurate predictions. However, it is crucial to interpret these results within a broader evaluative framework. Beyond accuracy and loss, factors such as the model's generalizability to unseen data, its performance across various patient subgroups, and a balanced consideration of precision and recall metrics are imperative for a holistic assessment of its clinical applicability and utility.

6. Conclusion

The research encapsulates a significant stride in the application of advanced machine learning techniques to pediatric healthcare, particularly in diagnosing and managing appendicitis. The multi-output neural network model developed and assessed in this study exhibits high accuracy and low loss in predicting management, severity, and diagnosis of pediatric appendicitis, underscoring its potential as a supplemental tool for clinicians. These findings represent a notable advancement in the integration of AI in medicine, providing a basis for further research and development of clinical decision support systems. The study not only contributes to the growing body of knowledge in medical machine learning but also opens avenues for future research, especially in the validation and generalization of such models in diverse clinical settings. As machine learning continues to evolve, its integration into healthcare promises to enhance diagnostic accuracy, personalize treatment strategies, and ultimately improve patient outcomes in pediatric appendicitis and beyond.

REFERENCES

[1] Andersson, Roland E. "The natural history and traditional management of appendicitis revisited: spontaneous resolution and predominance of prehospital perforations imply that a correct diagnosis is more important than an early diagnosis." World journal of surgery 31 (2007): 86-92.

[2] Marcinkevics, Ricards, et al. "Using machine learning to predict the diagnosis, management and severity of pediatric appendicitis." Frontiers in Pediatrics (2021): 360.

[3] Acharya, Amish, et al. "Biomarkers of acute appendicitis: systematic review and cost–benefit trade-off analysis." Surgical endoscopy 31 (2017): 1022-1031.

[4] Shommu, Nusrat S., et al. "The use of metabolomics and inflammatory mediator profiling provides a novel approach to identifying pediatric appendicitis in the emergency department." Scientific reports 8.1 (2018): 4083.

[5] Svensson, Jan F., et al. "Nonoperative treatment with antibiotics versus surgery for acute nonperforated appendicitis in children: a pilot randomized controlled trial." (2015): 67-71.

[6] Svensson, J. F., et al. "A review of conservative treatment of acute appendicitis." European Journal of Pediatric Surgery 22.03 (2012): 185-194.

[7] İlyas, Ö. Z. E. R. "Uzun kısa dönem bellek ağlarını kullanarak erken aşama diyabet tahmini." Mühendislik Bilimleri ve Araştırmaları Dergisi 2.2 (2020): 50-57.

[8] Ozer, Ilyas, et al. "Improved machine learning performances with transfer learning to predicting need for hospitalization in arboviral infections against the small dataset." Neural Computing and Applications 33 (2021): 14975-14989.

[9] Bayram, Muhammed Ali, Ö. Z. E. R. İlyas, and Feyzullah Temurtaş. "Deep learning methods for autism spectrum disorder diagnosis based on fMRI images." Sakarya University Journal of Computer and Information Sciences 4.1 (2021): 142-155.

[10] İlyas, Ö. Z. E. R. "Recurrent neural network based methods for hepatitis diagnosis." International Symposium of Scientific Research and Innovative Studies. Vol. 22. 2021.

[11] Temurtas, F., Gorur, K., Cetin, O., & Ozer, I. (2023). Machine learning for thyroid cancer diagnosis. In Computational Intelligence in Cancer Diagnosis (pp. 117-145). Academic Press.

[12] Marcinkevičs, Ričards (2023). Auto MPG. UCI Machine Learning Repository. https://doi.org/10.24432/C5859H.