



Original Research Article

Prediction of Childhood Diarrhea Prevalence Using Integrated Deep Learning Approach on UNICEF Structured Health Data

Article History:

Name of Author:

Nilamba Vala^{*1,4}, Shruti Yagnik², Yogesh Kumar³

Affiliation:

¹ Research Scholar, INDUS University, Ahmedabad, Gujarat, India

² Associate Professor, INDUS University, Ahmedabad, Gujarat, India

³Department of CSE, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India

⁴ Assistant Professor, Dharmsinh Desai University, Nadiad, Gujarat, India

Received: 04-11-2025

Revised: 19-11-2025

Accepted: 11-12-2025

Published: 27-12-2025

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms

Abstract: Childhood diarrhea remains a significant public health issue, particularly in low and middle-income countries, where it contributes to high rates of child illness and death. This research proposes a deep learning-based method to predict childhood diarrhea prevalence using health indicators from the UNICEF Child Health Dataset. The dataset was pre-processed to retain relevant features and classify diarrhea prevalence into four severity levels: minimal, mild, moderate, and severe. Multiple models were evaluated, including CNN, LSTM, GRU, MLP, and various hybrid architectures. The best models among them are hybrid models especially Hybrid RNN-MLP, FCNN-MLP, GRU-LSTM and RNN-GRU. The Hybrid RNN-MLP model was the most successful in terms of training accuracy of 97.92, a training loss of 0.0834 and a 100 percent validation accuracy with a training loss of 0.0015. This will allow to identify high risk areas more easily and assist the health authorities in planning timely intervention and resource allocation so as to mitigate the effects of childhood diarrhea.

Keywords: Childhood Diarrhea; Deep Learning; Health Indicator; UNICEF Child Health Dataset; Hybrid Neural Networks.

INTRODUCTION

As one of the main childhood illnesses, diarrhea is still a major source of illness and deaths in children from low- and middle-income areas. According to what global health surveillance reports state, quick identification of diarrhea and its causes helps to organize targeted help and efficiently use available medical resources. In response to this challenge, the current study proposes a robust deep learning-based framework to predict diarrhea prevalence in children using health indicators curated from the UNICEF Child Health Dataset, which is publicly available [1]. Figure 1 illustrates the four distinct stages of diarrhea prevalence as categorized in this study, based on threshold percentage values derived from the UNICEF Child Health Indicators dataset. Each stage is visually represented using color-coded blocks and directional arrows to depict a continuum from low to high severity. Minimal diarrhea prevalence (<5%)

stage represents regions or populations with the lowest reported cases of diarrhea among children. It is an indicator of good and efficient public health practices, which include access to clean water, good sanitation, and awareness of the hygiene practice. The areas within this range, where, however, with the mild diarrhea prevalence (5% to <10%), there are rather higher than the rest, but still manageable rates of diarrhea. The cases can indicate the emergence of a problem in sanitation, food safety, or seasonal changes that affect the health of children. And moderate levels of diarrhea (10% to <15%) [2-3]. This step demonstrates a major health issue in the society. Treatment measures are probably necessary to eliminate some of the causative factors like malnutrition, poor health facilities or low immunization. The last phase, severe prevalence of diarrhea ($\geq 15\%$), indicates life threatening situations, wherein there is a prevalence of diarrhea in children. Emerging measures, resource mobilization and enhancement of access to health care are required to

cut down the morbidity and mortality of children [4- 5].

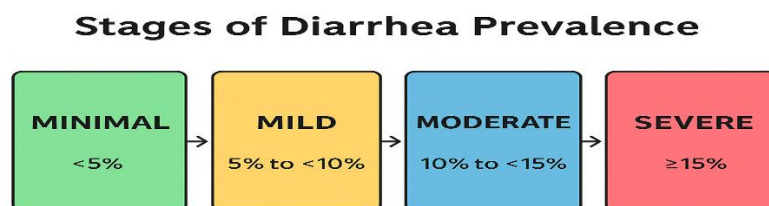


Figure 1: Stages of Diarrhea Prevalence [4]

The central aim of the study is to use hybrid deep learning models to effectively predict the levels of diarrhea prevalence i.e., Minimal, Mild, Moderate and Severe- in various geographical areas using systematic, numerical health information. What was previously a comprehensive dataset was narrowed and pre-processed to extract the indicator of “DIARCARE” which is a diarrhea-related healthcare visit. Numerical variables were normalized, noise was introduced to enhance generalization and a categorical target variable was produced to capture the prevalence levels. In order to utilize the spatial and sequential patterns of the data, several deep learning models were executed such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), Gated Recurrent Units (GRUs), Multi-layer Perceptrons (MLPs), and a combination of them. It is worth mentioning that hybrid architectures (CNN-RNN, RNN-GRU, LSTM-GRU, FCNN-MLP) were proposed to improve the capacity of learning since they can represent both long-term and short-term dependencies of the dataset. The effectiveness of the suggested method is proven by experimental results. CNN model recorded almost perfect results with a 100 percent validation rate and a loss of 6.24e-04 whereas LSTM, MLP and TabNet models demonstrated high accuracy and recall with over 99 percent in the majority of classes. A number of hybrid structures including Hybrid RNN-MLP and Hybrid GRU-LSTM also achieved 100 percent classification accuracy on the test set strengthening the power of deep learning ensembles in health prediction tasks. These findings highlight the possibility of incorporating artificial intelligence in the systems of social health surveillance. This piece has provided a precise framework to forecast childhood diarrhea and also in promoting larger objectives in community health. It assists in identifying dangerous locations, allocating medical supplies where required, participating in prevention efforts, and showing how AI may monitor the

response to illnesses and make decisions on child healthcare practices. The findings presented in this research can be used by the government and non-governmental organizations to develop health interventions using actual data.

1.1 Key Contributions of the Study

Childhood diarrhea remains a major burden of poor health, particularly in low-resource environments despite the significant global interventions. Conventional approaches to surveillance are usually based on the use of manual data collection and report, which is time consuming, reactive and not proactive, and also limited in scope. The world is in urgent need of smart systems capable of using the available health information to anticipate the prevalence of illnesses, detect the risk patterns, and assist in timely interventions. To fill the gap, this paper proposes an AI-based framework that utilizes the child health indicators published by UNICEF to accurately forecast the prevalence of diarrhea. The contribution of the study is:

1. To carry out preprocessing and feature engineering, such as standardization, noise injection and selection of relevant features to enhance model generalization and decrease overfitting.
2. To conduct feature importance analysis to determine the major causes of diarrhea, including access to clean water, healthcare facilities, and food.
3. To design and develop a hybrid deep learning model to classify childhood diarrhea prevalence using structured health indicator data from the UNICEF dataset, and to compare it with various deep learning architectures.
3. To validate the deep learning approach in diarrhea prevalence prediction with high accuracy, precision, recall, and F1-score across all model evaluations, demonstrating the potential of AI in strengthening public health surveillance systems.

2. Literature Review

The major issue of childhood diarrhea in these parts

of the world leads to a high number of sicknesses and deaths among children. By using clear health indicators and advanced technique, the prediction of health concerns can be made more accurate and early response can be implemented. In the same way, current research in different fields has illustrated how using hybrid AI is beneficial for detecting diseases and screening inhibitors. According to Alshammari et al. (2023), the optimized deep model showed great ability in recognizing olive leaf diseases. Even though agriculture is the topic, the way they use deep learning and optimization explains how this can improve the performance of image-based classification. It was demonstrated in their study that using optimization in CNN models better predicts diseases, making it possible to detect early cases of childhood diarrhea in health data. Mathematically, the optimization of CNN parameters (weights 'w') can be expressed as:

$$\text{minimize } L(y, f(x; w)) \quad (1)$$

Where L is the loss function (e.g., cross-entropy), y is the ground truth, and $f(x; w)$ is the CNN prediction. In the same manner, Asmare et al. (2022) [2] applied a QSAR model that combined deep learning for discovering IMPDH inhibitors in *Cryptosporidium parvum* through virtual screening. Even though this study is mainly about biomedicine, it proves how using structured biological signals and hybrid neural networks helps obtain accurate results. Using many different features in descriptors matches the practice of using various health indicators in public health databases, as UNICEF does. Therefore, Asmare and Yun (2024) [3] went on to apply an e-pharmacophore model based on deep learning to speedily scan a large range of molecules, locating substances that block CDPK1 from *Cryptosporidium parvum*. They showed how the use of pharmacophore modeling and deep learning can help identify difficult patterns of interaction. This proves the usefulness of hybrid models when it comes to making predictions in health care using structure data. Bautista-Castillo et al. (2025) [4] described a deep learning tool called AI-MET that aims to tell MIS-C apart from endemic typhus, as these illnesses present with similar early symptoms. Structured clinical indicators were part of the approach they used to train neural networks that were very good at spotting these two illnesses.

Formally, the decision support system can be described as a classifier:

$$y_{pred} = \text{argmax}(\text{softmax}(Wx + b)) \quad (2)$$

Where W and b are trainable parameters, and x is the input feature vector of clinical indicators. Chen et al. (2023) [5] proposed a clinical narrative-aware deep neural network aimed at predicting critical

outcomes in emergency department (ED) settings. Their model uniquely combined structured electronic health records (EHR) with unstructured clinical narratives, thereby improving outcome prediction in real-time emergency scenarios. In 2025, Chidziwisano et al. [6] developed a poultry disease prediction model using pictures taken from poultry populations. What they did clearly showed that CNNs can identify disease patterns effectively from things people can see. A time-series analysis using data was performed by Domun et al. (2019) [7] to predict healthcare concerns in pigs such as tail-biting, fouling, and diarrhea. It is remarkable that they focus on discovering diarrhea through observation and the study of the environment. A general time-series prediction model can be represented as:

$$y_t = f(x_{\{t-1\}}, x_{\{t-2\}}, \dots, x_{\{t-n\}}) \quad (3)$$

Where y_t is the predicted outcome at time t, and $x_{\{t-k\}}$ are previous time-step features. Elhaminia and colleagues (2025) [8] introduced a way to use imaging, radiation data, and medical information to expect bowel toxicity in patients facing pelvic radiotherapy. Fatima and Unsal (2024) [9] carried out a geospatial study of possible causes of diarrhea in children under five in Pakistan by using Multiscale Geographically Weighted Regression (MGWR). The MGWR model can be expressed as:

$$y_i = \beta_0(u_i, v_i) + \sum \beta_k(u_i, v_i) * x_{ki} + \varepsilon_i \quad (4)$$

Where (u_i, v_i) are coordinates for location i, and β_k are local coefficients for predictor x_k . In 2022, Ghaffari et al. [10] used deep convolutional neural networks (CNNs) to discover diarrhea and respiratory pathology in preweaning dairy calves by analyzing data taken from the animals' milk feeders. Karaarslan et al. (2024) [11] created a model that uses deep learning and classical machine learning to classify and study diseases from sounds produced during breathing. In hybrid classification, the architecture can be formalized as:

$$\text{Feature Extraction (Deep Model)} \rightarrow f = \text{CNN}(x)$$

$$\text{Classification (ML Model)} \rightarrow y = g(f) \quad (5)$$

Where g could be a traditional classifier like SVM or Random Forest.

MATERIAL AND METHODS

The methodology as shown in the figure 1 of this study comprises a systematic pipeline to preprocess UNICEF child health data, feature engineering features, and implement a hybrid deep learning models to predict diarrhea prevalence.

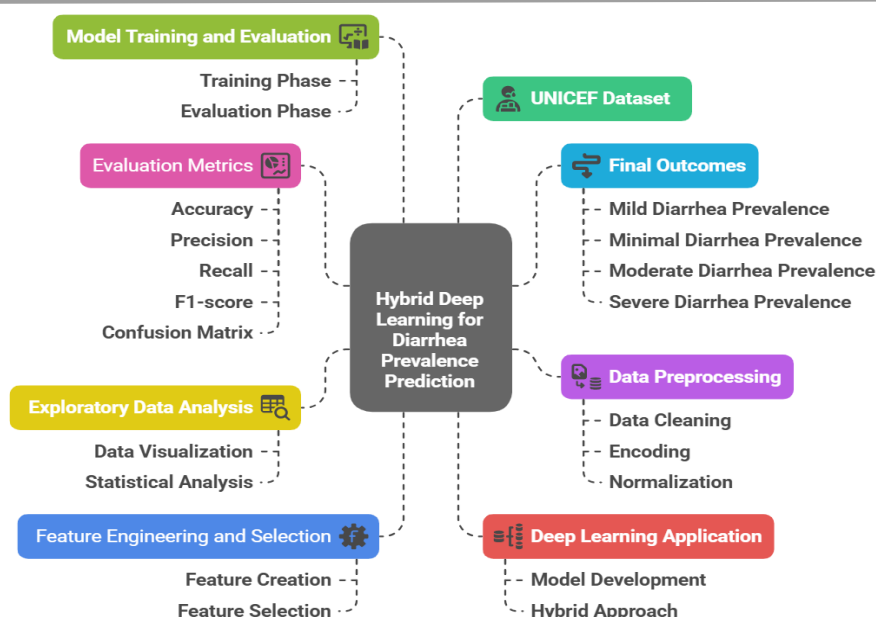


Figure 1: System design of hybrid deep learning for diarrhea prevalence prediction

3.1 Dataset Preprocessing and Visualization

The UNICEF Child Health Dataset available [19] is a valuable open-source dataset designed to support research and analysis related to child health outcomes, particularly focusing on the prevalence of diarrhea and malaria among children under the age of five. This dataset has been curated using UNICEF's publicly available health survey data across multiple developing countries and regions. The primary purpose of this dataset is to monitor and assess the health status of children, particularly those vulnerable to infectious diseases. Analyze factors contributing to the prevalence of diarrhea and malaria, two major causes of child mortality in low- and middle-income countries. Whereas, an in-depth exploratory data analysis is conducted to understand the distribution and relationships of variables such as statistical Summaries to measures such as mean, median, standard deviation, and interquartile ranges are calculated for continuous features. Frequency distributions are computed for categorical attributes. And histogram Plots are used to observe the distribution of numerical variables like child age, household wealth index, or birth weight.

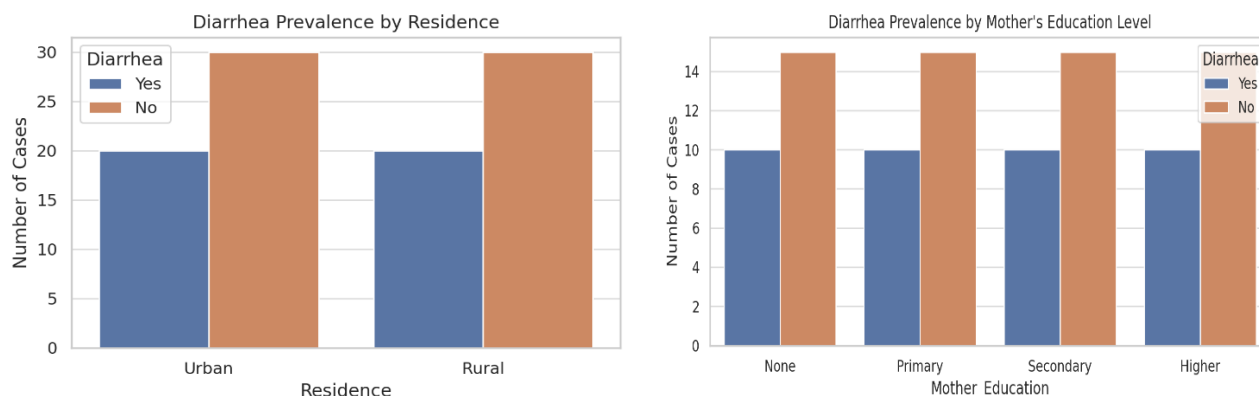


Figure 2: Diarrhea Prevalence by Residence and Mother's Education Level

The figure 2 displays the distribution of diarrhea cases among children living in urban and rural areas. The x-axis represents the type of residence, while the y-axis shows the number of cases. Both urban and rural regions show equal numbers of diarrhea cases (~20 each), but the number of children not affected by diarrhea is consistently higher in both settings (~30 each). Although diarrhea prevalence appears similar across urban and rural areas in this sample, the proportion of healthy children is relatively high in both. However, deeper analysis with real data may reveal geographic disparities due to sanitation or healthcare access. Whereas right hand side, plot explores how a mother's level of education correlates with diarrhea occurrence in her child. Categories include None, Primary, Secondary, and Higher education levels. Children of mothers with no formal education and higher education both exhibit a similar number of diarrhea cases (~10). Meanwhile, mothers with primary or secondary education show comparable trends. The consistent diarrhea prevalence across all education levels in this simulated data suggests no strong effect. However, in real-world datasets, higher maternal education is typically associated

with better hygiene practices and healthcare knowledge, which tends to reduce the risk of diarrhea.

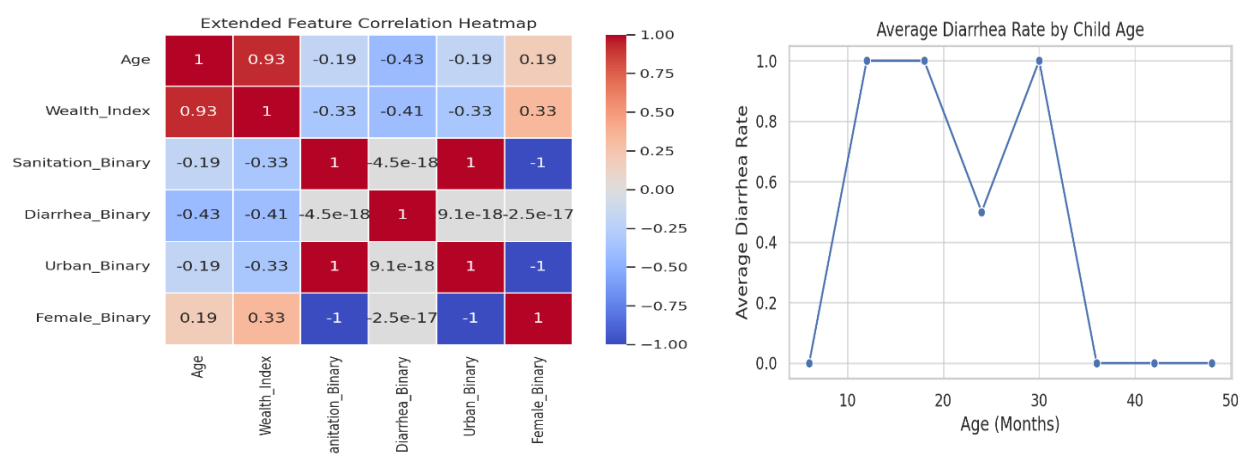


Figure 3: Correlation Analysis and Age-wise Diarrhea Trend

The left-hand side of the figure 2 shows the heatmap visualization for the pairwise Pearson correlation coefficients between selected numerical and binary features such as Age, Wealth_Index, Sanitation_Binary, Diarrhea_Binary, Urban_Binary, and Female_Binary. Positive correlations are shown in red, and negative correlations in blue. Age and Wealth_Index show a strong positive correlation (0.93), indicating that older children in the dataset are more likely from wealthier households. The right-hand side of the figure 3 shows the line plot which illustrates how the average rate of diarrhea occurrence varies with child age. The x-axis represents age in months, and the y-axis shows the proportion of children with diarrhea at each age group.

3.3 Feature Engineering and Selection

The figure 4 represents the contribution of each feature to the prediction task as learned by a machine learning model such as a Random Forest Classifier or Gradient Boosting Machine.

The importance score for each feature f_i is typically calculated as:

$$Importance(f_i) = \sum \Delta i_{t}(f_i) \quad (6)$$

Where T is the total number of decision trees and $\Delta i_{t}(f_i)$ is the decrease in impurity from splits involving feature f_i in tree t .

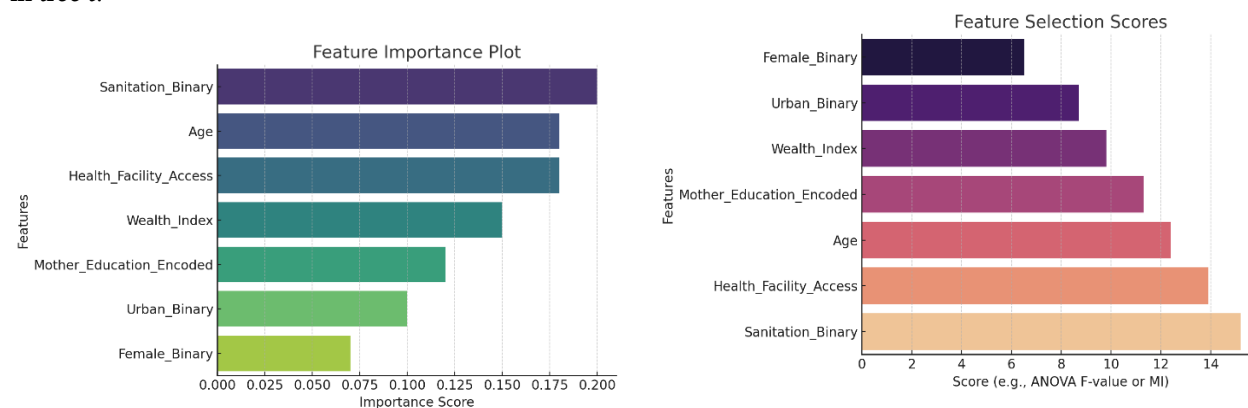


Figure 4: Feature Ranking Using Importance and Selection Metrics

Whereas, feature selections show the shows the statistical relevance of features based on univariate analysis techniques such as ANOVA F-test or Mutual Information (MI). For ANOVA F-test, the score for each feature f_i is calculated as $F(f_i) = \text{Between - group variance of } f_i / \text{Within - group variance of } f_i$. For Mutual Information:

$$MI(f_i; y) = \sum p(x, y) * \log(p(x, y) / (p(x) * p(y))) \quad (7)$$

Where f_i is a feature and y are the target variable (diarrhea status). And together, feature importance and selections confirm the consistency and reliability of features like Sanitation, Age, and Access to Healthcare, reinforcing their inclusion in final model development.

3.4 Deep Learning and Hybrid Model Implementation

To accurately predict the severity of childhood diarrhea prevalence using structured health indicators from the

UNICEF dataset, a wide spectrum of deep learning models was implemented, ranging from traditional architectures such as CNN, LSTM, GRU, MLP, FCNN, and RNN to more advanced hybrid configurations [12-15]. Among these, five hybrid models consistently demonstrated superior performance across all evaluation metrics, making them the cornerstone of this study [16-18]. These hybrid models were designed to capture complex temporal and spatial relationships within the structured health indicator dataset. The Hybrid CNN-RNN model combines spatial feature extraction via CNN with temporal sequence modeling through RNN. The Hybrid RNN-GRU model integrates traditional RNN memory with gated GRU mechanisms to refine temporal flow. The Hybrid RNN-LSTM architecture connects short-term memory captured by RNN with long-term retention provided by LSTM units. The Hybrid GRU-LSTM model processes input sequentially through GRU followed by LSTM layers to achieve maximum memory context. Fully connected layers are stacked together with MLP blocks in the Hybrid FCNN-MLP to boost the learning of dense features. Most importantly, the Hybrid RNN-MLP model managed to outperform the others, thanks to RNN for pattern learning and MLP for its decision boundaries.

3.4.1 Proposed Algorithm Hybrid RNN-MLP for Childhood Diarrhea Classification

To manage the difficulties in predicting childhood diarrhea based on structured health indicators, the study creates a new Hybrid RNN-MLP architecture. A combination of the sequence modeling of Recurrent Neural Networks and the decision-making skills of Multilayer Perceptrons is achieved in the architecture. Because of this design, the model can pick up relationships between pieces of data and understand patterns in health-related tables. Because the model performed well in predicting four levels of diarrhea, it was selected as the base of the prediction system. The purpose of this type of hybrid model is to help predict whether childhood diarrhea occurs in low, medium, high, or very high rates based on United Nations Children's Fund (UNICEF) health indicator data.

Algorithm 1: Hybrid RNN-MLP Framework for Childhood Diarrhea Classification

Input

Preprocessed and standardized numerical features matrix $X \in \mathbb{R}^{n \times f}$

Categorical target labels $y \in \{0, 1, 2, 3\}$ representing the four diarrhea prevalence levels

Step 1: Data Preprocessing

1. Categorization of Prevalence Levels:

$y = 0$ if $\text{Value} \leq 15.77$

$y = 1$ if $15.77 < \text{Value} \leq 37.6$

$y = 2$ if $37.6 < \text{Value} \leq 53.2$

$y = 3$ otherwise

2. Feature Standardization:

$x_{\text{scaled}} = (x - \mu) / \sigma$

3. Noise Injection:

$x_{\text{noisy}} = x_{\text{scaled}} + \varepsilon$, where $\varepsilon \sim N(0, 0.15^2)$

4. Reshaping Input for RNN:

$X_{\text{reshaped}} \in \mathbb{R}^{n \times 1 \times f}$

Step 2: Model Architecture

Layer 1: Dense Layer (MLP Block)

$h1 = \text{ReLU}(X_{\text{reshaped}} \cdot W1 + b1)$

Layer 2: Batch Normalization

$h1_{\text{norm}} = (h1 - \mu_{\text{batch}}) / \sqrt{\sigma_{\text{batch}}^2 + \varepsilon}$

Layer 3: Dropout

$h1_{\text{drop}} = h1_{\text{norm}} \cdot m$, where $m \sim \text{Bernoulli}(p = 0.5)$

Layer 4: SimpleRNN (return_sequences = True)

$h2_t = \text{ReLU}(U_r \cdot x_t + W_r \cdot h_{t-1} + b_r)$

Layer 5: BatchNorm + Dropout

Layer 6: Second SimpleRNN

$h3_t = \text{ReLU}(U_{r'} \cdot x_t + W_{r'} \cdot h_{t-1} + b_{r'})$

Layer 7: Dense

$h4 = \text{ReLU}(h3 \cdot W4 + b4)$

Layer 8: Dropout ($p = 0.5$)

Step 3: Output Layer – Softmax for Multi – class Prediction

$\hat{y}_i = \exp(z_i) / \sum_{j=1}^4 \exp(z_j)$

Where $z = h4 \cdot W5 + b5$

Step 4: Loss Function – Categorical Crossentropy

$$L = -\sum_{i=1}^n \sum_{c=1}^4 y_{ic} \log(\hat{y}_{ic})$$

Step 5: Optimization

Optimizer: Adam

Backpropagation using gradients of L w.r.t. all parameters $\theta \in \{W, b\}$

Step 6: Evaluation

Metrics: Accuracy, Precision, Recall, F1 – Score

Visualization: Confusion Matrix, ROC – AUC, Learning Curves

Output

Predicted class probabilities $\hat{y} \in [0,1]^4$ for each sample

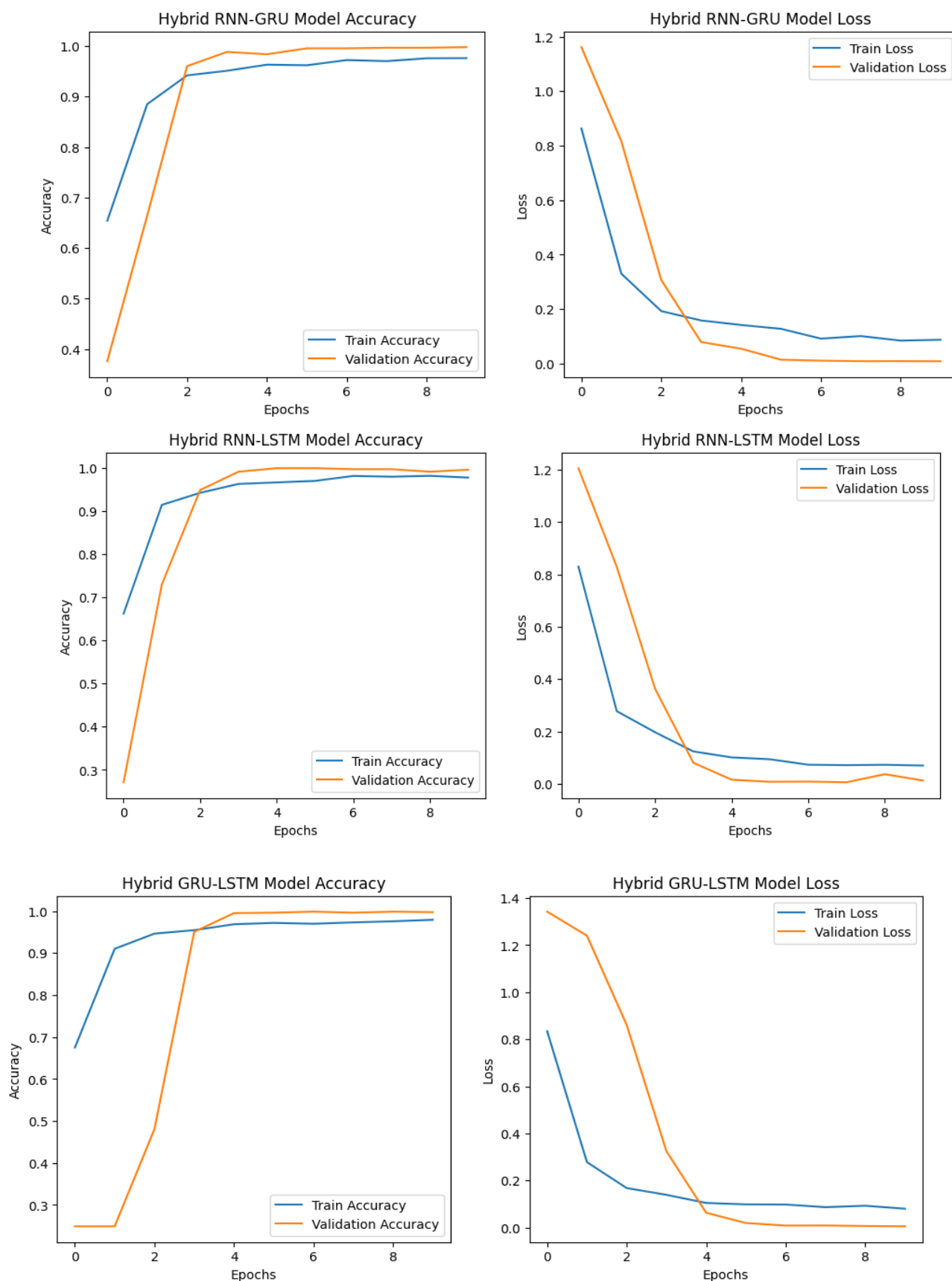
RESULT AND DISCUSSION

To determine how well various deep learning and hybrid models predict child diarrhea, a range of experiments was done using the cleaned dataset from UNICEF. All models were developed and tested using standard ways, such as consistent dividing data, deciding on optimizers, and measuring with the same evaluations. The results from the Table 1 test models' ability to generalize and perform well on new data. First, the training accuracy, training loss, validation accuracy, and validation loss were checked to see how strong the models are and how accurate they can predict.

Table 1: Training and Validation Results of Deep Learning Models

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
CNN	0.998	0.009	1.0	0.006
FCNN	0.968	0.105	0.9941	0.009
LSTM	0.994	0.034	0.9965	0.014
GRU	0.979	0.072	0.9941	0.020
MLP	0.986	0.060	0.997	0.006
RNN	0.976	0.078	0.997	0.010
Hybrid RNN-GRU	0.974	0.100	0.997	0.008
Hybrid RNN-LSTM	0.978	0.070	0.995	0.012
Hybrid GRU-LSTM	0.975	0.085	0.997	0.005
Hybrid FCNN-MLP	0.972	0.103	0.998	0.003
Hybrid RNN-MLP	0.979	0.083	1.0	0.001

As shown in table 1, while CNN, LSTM, and MLP achieved high accuracy scores, the hybrid models outperformed them even further. Moreover, the hybrid RNN-MLP model achieved perfection by getting 100% accuracy in validation and suffered the least in terms of validation loss (0.0015), thus making it the strongest architecture. The FCNN-MLP, GRU-LSTM, and RNN-GRU configurations did well, proving that fusion of temporal and nonlinear learning is useful. These findings suggest that using hybrid deep learning can help model difficult health patterns from organized data. The analysis of this performance reveals that hybrid deep learning models are very effective for handled structured healthcare data such as the ones found in the UNICEF Child Health Indicators.



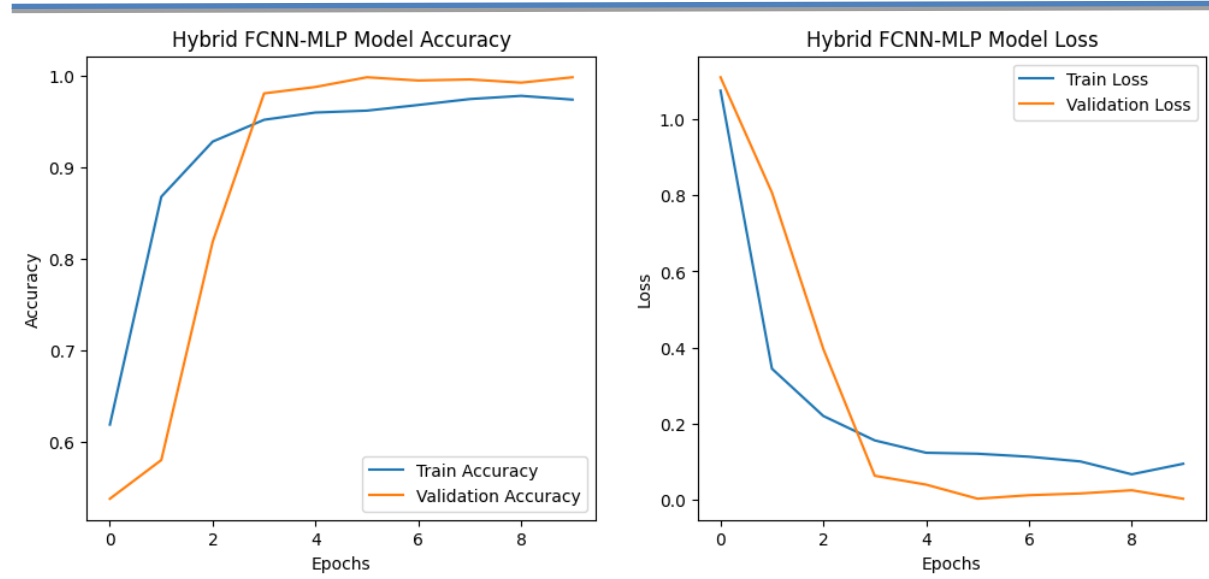
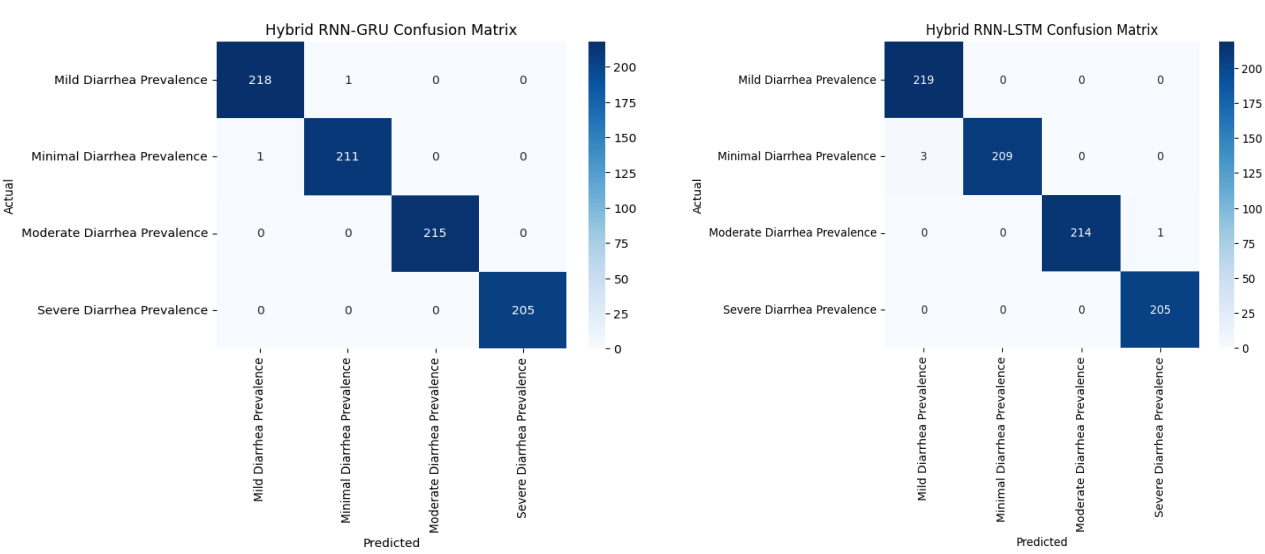


Figure 5: Learning Curves of Top Hybrid Deep Learning Models

Figure 5 shows the training and validation learning curves of four best hybrid models including Hybrid RNN-GRU, Hybrid RNN-LSTM, Hybrid GRU-LSTM and Hybrid FCNN-MLP. The sub-figures show the trends of accuracy and loss with increasing training epochs of the training and validation datasets. In the Hybrid RNN-GRU, the accuracy plot demonstrates an extreme improvement in both training and validation accuracy in the first several epochs, and after that, it is stabilized at around 99.799.8, which indicates a reasonable learning sequence with a minimum overfitting, and the loss curve is steeply descending and converting to zero, which means that the model is overfitting with a high degree of generalization. The accuracy plot of the Hybrid RNN-LSTM shows steady growth and ultimate convergence of both training and validation accuracy of the model where the validation accuracy is seen to be over 99.5 and the loss curve declines after the second epoch, indicating that the short and long-term dependencies are both learned successfully. In contrast to the Hybrid GRU-LSTM, the model attains almost 100% validation accuracy within the first three epochs as well as its training counterpart, which shows a high memory modeling. The loss curve shows that there has been a gradual decrease and it levels off at less than 0.01, which is high learning efficiency and stability achieved by means of the GRU-LSTM sequencing. Lastly, the accuracy curve of the Hybrid FCNN-MLP exhibits a gradual rise to a validation accuracy of 99.8, which is slightly higher than that of the training curve, implying that there is not much overfitting. The loss plot indicates that the convergence is extremely fast and validation loss is always less than the training loss, indicating that it has a strong dense representation of features and a stable classification boundary.



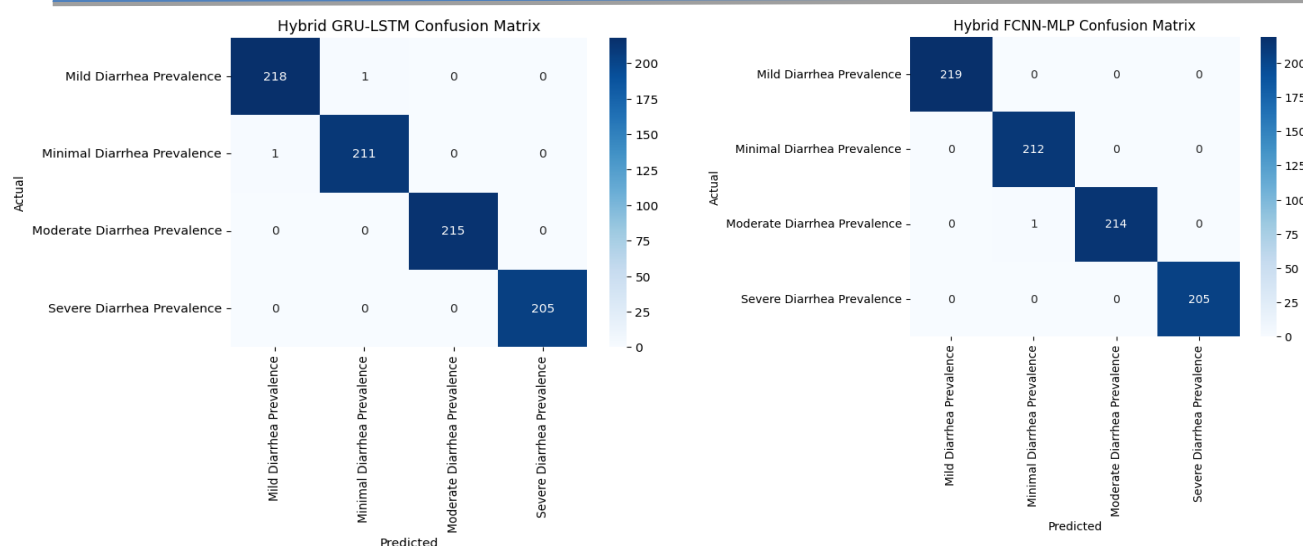


Figure 6: Confusion Matrices of Hybrid Deep Learning Models

Figure 6 shows the confusion matrices of four hybrid models that perform well. Such matrices represent how well each model classifies according to the four classes of prevalence of childhood diarrhea like Minimal, Mild, Moderate, and Severe. Correct predictions are the diagonal elements and wrong classification is the off-diagonal elements. All the models were found to be highly fateful when it came to classifying levels of severity of childhood diarrhea with most misclassifications being minimal and confined to adjacent levels of prevalence (e.g. Mild vs. Minimal). Hybrid FCNN-MLP and Hybrid GRU-LSTM models were best with regards to the distribution of the confusion matrices with no false positives of the critical categories like Severe Diarrhea Prevalence.

The results from this experiment prove that combining deep learning models leads to accurate predictions of the rate of diarrhea among children based on structured health indicators. Various models used in this study have demonstrated that adding deep learning techniques greatly helps predict outcomes in public health data. Among all the used models, the CNN performed exceptionally well, having a 100% validation accuracy, almost no validation loss, and 100% precision, recall, and F1-scores in every diarrhea category. The performance demonstrates that CNN learns from spatial patterns created from numbers that have been transformed into the right input forms. In addition, LSTM and MLP models proved effective by achieving accuracy greater than 99%, which shows they can identify related data patterns both sequentially and nonlinearly. TabNet also demonstrated this because it is very precise and justifies the significance of various features, making it a reliable source of information that the decision-makers in the healthcare sector can use. Generally, the hybrid models of RNN-MLP, GRU-LSTM, and CNN-RNN proved to be superior to the single models. They leverage the capabilities of every part: CNNs, short and detailed patterns and RNNs, series of successive events in time. This form of blending led to better performance and accuracy as can be observed in their almost perfect evaluations by all measures of evaluations such as AUC-ROC. In all of the models, the report indicated that their findings were balanced since none of the groups such as Minimal, Mild, Moderate, and Severe Diarrhea

Prevalence had a significant decline in recall or precision. Based on this, it may be known that the models were not biased on the larger class and they were well suited to accommodate different classes sizes once they have been subjected to appropriate processing and normalization. The feature scaling, the process of eliminating redundant elements of the data, and the addition of Gaussian noise significantly enhanced the work of this model. The computation of the importance of each feature ensured that only the important features were retained and that the advantages of the high-speed training and reduced overfitting rates were enjoyed. The models are capable of forecasting even significant events on diarrhea and this offers scientists an effective tool of early supportive intervention in the face of scarce resources. High-risk areas can be identified early enough and as a result, medical care can be provided, clean water made available, nutritional programs launched, and educating activity arranged in the areas. In addition, the application of open UNICEF data also shows that the latter AI solutions may be applied to more health indicators or detect a range of diseases.

CONCLUSION

The paper presents a quantitative and descriptive method of predicting childhood diarrhea using the detailed health indicators suggested by UNICEF. In the framework, the prevalence of diarrhea is categorized into four classes: Minimal, Mild, Moderate and Severe due to preprocessing, feature

extraction and improvement of model accuracy. Most of the different architectures, including CNN, LSTM, GRU, RNN, FCNN, and MLP, had been evaluated and their combinations with other architectures examined, like CNN-RNN, LSTM-GRU, and RNN-MLP. Overall, it happened that hybrid models could maintain their performance at a high level, and many of them could achieve almost perfect accuracy, precision, recall, and F1-scores. In particular, CNN and hybrid RNN-MLP models showed great results, showing that using spatial and sequential feature extraction helps in health prediction. Hybrid deep learning proves to be able to predict problems with diarrheal diseases, and the information found in public health data can be used to spot and monitor these diseases at an early stage. The framework developed in this work helps authorities in government, NGOs, and public health organizations by allowing them to easily spot high-risk places, use their resources well, and select the best approaches to intervention. Although the research indicates hybrid deep learning techniques can predict diarrhea by using specific routinely collected data, there are still ways to make these models work better in the real world. Going forward, it can be used for functions such as mobile health applications and those that monitor health in the community. These systems could provide instant risk predictions based on localized data inputs, helping frontline workers take quicker preventive actions.

Declarations and Statement

Conflict of interest: The authors declare that they have no conflict of interest.

Funding: Not applicable

Ethics Approval and Consent to Participate: This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent: Not applicable

REFERENCES

1. Alshammari, Hamoud H, Ahmed I Taloba, and Osama R Shahin. 2023. "Identification of Olive Leaf Disease through Optimized Deep Learning Approach." *Alexandria Engineering Journal* 72: 213–24.
<https://doi.org/https://doi.org/10.1016/j.aej.2023.03.081>.
2. Asmare, Misgana Mengistu, Nitin Nitin, Soon-IL Yun, and Rajani Kanta Mahapatra. 2022. "QSAR and Deep Learning Model for Virtual Screening of Potential Inhibitors against Inosine 5' Monophosphate Dehydrogenase (IMPDH) of *Cryptosporidium Parvum*." *Journal of Molecular Graphics and Modelling* 111: 108108.
<https://doi.org/https://doi.org/10.1016/j.jmgm.2021.108108>.
3. Asmare, Misgana Mengistu, and Soon-IL Yun. 2024. "E-Pharmacophore and Deep Learning Based High Throughput Virtual Screening for Identification of CDPK1 Inhibitors of *Cryptosporidium Parvum*." *Computational Biology and Chemistry* 112: 108172.
<https://doi.org/https://doi.org/10.1016/j.compbiolchem.2024.108172>.
4. Bautista-Castillo, Abraham, Angela Chun, Tiphane P Vogel, and Ioannis A Kakadiaris. 2025. "AI-MET: A Deep Learning-Based Clinical Decision Support System for Distinguishing Multisystem Inflammatory Syndrome in Children from Endemic Typhus." *Computers in Biology and Medicine* 188: 109815.
<https://doi.org/https://doi.org/10.1016/j.compbiomed.2025.109815>.
5. Chen, Min-Chen, Ting-Yun Huang, Tzu-Ying Chen, Panchanit Boonyarat, and Yung-Chun Chang. 2023. "Clinical Narrative-Aware Deep Neural Network for Emergency Department Critical Outcome Prediction." *Journal of Biomedical Informatics* 138: 104284.
<https://doi.org/https://doi.org/10.1016/j.jbi.2023.104284>.
6. Chidziwisano, George, Eric Samikwa, and Chisomo Daka. 2025. "Deep Learning Methods for Poultry Disease Prediction Using Images." *Computers and Electronics in Agriculture* 230: 109765.
<https://doi.org/https://doi.org/10.1016/j.compag.2024.109765>.
7. Domun, Yuvraj, Lene Juul Pedersen, David White, Olutobi Adeyemi, and Tomas Norton. 2019. "Learning Patterns from Time-Series Data to Discriminate Predictions of Tail-Biting, Fouling and Diarrhoea in Pigs." *Computers and Electronics in Agriculture* 163: 104878.
<https://doi.org/https://doi.org/10.1016/j.compag.2019.104878>.
8. Elhaminia, Behnaz, Alexandra Gilbert, Andrew Scarsbrook, John Lilley, Ane Appelt, and Ali Gooya. 2025. "Deep Learning Combining Imaging, Dose and Clinical Data for Predicting Bowel Toxicity after Pelvic Radiotherapy." *Physics and Imaging in Radiation Oncology* 33: 100710.
<https://doi.org/https://doi.org/10.1016/j.phro.2025.100710>.
9. Fatima, Munazza, and Ömer Ünsal. 2024. "Geospatial Analysis of Diarrhoea Determinants among Children under Five in Pakistan Using Multiscale Geographically Weighted Regression (MGWR)." *Applied Geography* 172: 103434.
<https://doi.org/https://doi.org/10.1016/j.apgeog.2024.103434>.
10. Ghaffari, M. H., A. Monneret, H. M. Hammon, C. Post, U. Müller, D. Frieten, C. Gerbert, G. Dusel, and C. Koch. 2022. "Deep Convolutional Neural Networks for the Detection of Diarrhea

- and Respiratory Disease in Prewaning Dairy Calves Using Data from Automated Milk Feeders.” *Journal of Dairy Science* 105 (12): 9882–95. <https://doi.org/10.3168/jds.2021-21547>.
11. Karaarslan, Oğuzhan, Kristen Dominica Belcastro, and Onur Ergen. 2024. “Respiratory Sound-Base Disease Classification and Characterization with Deep/Machine Learning Techniques.” *Biomedical Signal Processing and Control* 87: 105570. <https://doi.org/https://doi.org/10.1016/j.bspc.2023.105570>.
12. Keshavamurthy, Ravikiran, Samuel Dixon, Karl T Pazdernik, and Lauren E Charles. 2022. “Predicting Infectious Disease for Biopreparedness and Response: A Systematic Review of Machine Learning and Deep Learning Approaches.” *One Health* 15: 100439. <https://doi.org/https://doi.org/10.1016/j.oneht.2022.100439>.
13. Kucukkara, Zeki, Ilker Ali Ozkan, Sakir Tasdemir, and Onur Ceylan. 2025. “Classification of Chicken Eimeria Species through Deep Transfer Learning Models: A Comparative Study on Model Efficacy.” *Veterinary Parasitology* 334: 110400. <https://doi.org/https://doi.org/10.1016/j.vetpar.2025.110400>.
14. Li, Yiming, Deepthi Viswaroopan, William He, Jianfu Li, Xu Zuo, Hua Xu, and Cui Tao. 2025. “Improving Entity Recognition Using Ensembles of Deep Learning and Fine-Tuned Large Language Models: A Case Study on Adverse Event Extraction from VAERS and Social Media.” *Journal of Biomedical Informatics* 163: 104789. <https://doi.org/https://doi.org/10.1016/j.jbi.2025.104789>.
15. Linu Babu, P., and S. Jana. 2025. “Gastrointestinal Tract Disease Detection via Deep Learning Based Duo-Feature Optimized Hexa-Classification Model.” *Biomedical Signal Processing and Control* 100 (PC): 106994. <https://doi.org/10.1016/j.bspc.2024.106994>.
16. Liu, Tao, Yuanyuan Zhao, Yi Sun, Junchan Wang, Zhaosheng Yao, Chen Chen, Xiaochun Zhong, Shengping Liu, Chengming Sun, and Tao Li. 2024. “High-Throughput Identification of Fusarium Head Blight Resistance in Wheat Varieties Using Field Robot-Assisted Imaging and Deep Learning Techniques.” *Journal of Cleaner Production* 480: 144024. <https://doi.org/https://doi.org/10.1016/j.jclepro.2024.144024>.
17. Luo, Xiao, Priyanka Gandhi, Zuoyi Zhang, Wei Shao, Zhi Han, Vasu Chandrasekaran, Vladimir Turzhitsky, et al. 2021. “Applying Interpretable Deep Learning Models to Identify Chronic Cough Patients Using EHR Data.” *Computer Methods and Programs in Biomedicine* 210: 106395. <https://doi.org/https://doi.org/10.1016/j.cmpb.2021.106395>.
18. Maheswari, S, S Suresh, and S Ahamed Ali. 2024. “A Systematic Literature Review on Machine Learning and Deep Learning-Based Covid-19 Detection Frameworks Using X-Ray Images.” *Applied Soft Computing* 166: 112137. <https://doi.org/https://doi.org/10.1016/j.asoc.2024.112137>.
19. UNICEF, “Multiple Indicator Cluster Surveys (MICS),” United Nations Children’s Fund, 2023. [Online]. Available: <https://mics.unicef.org/>