Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, March : 2025

OPTIMIZED SYMPTOM-BASED DEEP LEARNING FRAMEWORK FOR MONKEYPOX DIAGNOSIS WITH LIME EXPLAINABILITY

Bhaskar Kantapalli, Assistant Professor, Dept. of CSE, Seshadri Rao Gudlavalleru Engineering College, India
Arshia Aamena, Chebrolu Yogavarshinee, Badugu Divya Teja, Dasari Teja Sri, UG student, Dept. of CSE, Seshadri Rao Gudlavalleru Engineering College, India

Abstract:
Monkeypox is an emerging zoonotic disease that has raised global health concerns due to its increasing transmission rates. Traditional diagnostic methods rely on laboratory testing, which can be time-consuming and inaccessible in resource-limited settings. This study presents an Optimized Deep Neural Framework (ODNF) to diagnose monkeypox based on clinical symptoms, leveraging deep learning for accurate and rapid classification. The research explores various machine learning models, including Random Forest, XG Boost, and Cat Boost, before implementing ODNF, which achieved superior performance with a 99% accuracy rate. The dataset underwent preprocessing steps, including handling imbalanced data and feature encoding, ensuring optimal learning. Additionally, Local Interpretable Model-Agnostic Explanations (LIME) was employed to enhance model interpretability, providing insights into symptom-based predictions. Comparative evaluation against traditional models demonstrated that ODNF outperforms existing approaches, making it a viable AI-based diagnostic tool for monkeypox detection.

Keywords: Optimized Deep Neural Framework, Monkeypox Diagnosis, AI-based Healthcare, LIME

1. Introduction

The 1970s saw the discovery of monkeypox, a viral zoonotic disease, in the Democratic Republic of the Congo [1]. Outbreaks have been documented globally since May 2022, and by August 2022, there were over 318,000 confirmed cases [2][3]. The virus, classified under the Poxviridae family, is closely related to smallpox and presents symptoms such as fever, skin lesions, lymphadenopathy [4], and respiratory complications. The rapid transmission of monkeypox through direct human-to-human contact, respiratory droplets, and contaminated surfaces has posed significant public health concerns [5].

Reducing the disease's spread requires prompt and accurate diagnosis. Antigen detection and polymerase chain reaction (PCR) testing are examples of traditional techniques that are impracticable in environments with low resources since they call for specialised labs and skilled workers [6]. Furthermore, symptom overlap with other infectious diseases often leads to misdiagnosis. Medical diagnostics have benefited greatly from the automated and scalable solutions provided by artificial intelligence (AI) and machine learning (ML) approaches. In recent years, deep learning approaches, particularly convolutional neural networks (CNNs), have been used for image-based monkeypox detection. However, CNN-based methods require high-quality image datasets and suffer from computational inefficiencies [7]. This study introduces the Optimized Deep Neural Framework (ODNF), a novel ANN-based approach that leverages clinical symptom data for early monkeypox detection. Unlike traditional methods, ODNF integrates advanced optimization strategies such as adaptive learning rate scheduling, dropout regularization, and batch normalization to improve classification performance. We evaluate various ML models, including Random Forest, XGBoost, and CatBoost, before demonstrating the superiority of ODNF. The model's interpretability is enhanced using LIME explainability, enabling medical practitioners to understand the significance of specific symptoms in classification decisions.

Our key contributions include:

1. Development of a deep learning-based ODNF model trained on clinical symptom data, achieving 99% accuracy.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, March : 2025

2 Comparative assessment of multiple machine learning models, including Random Forest and boosting algorithms .
3 The use of Local Interpretable Model-agnostic Explanations (LIME) to enhance model interpretability and support clinical judgement.
4 Scalability and real-time applicability, making ODNF a practical alternative to existing diagnostic tools.

2. Related Work

Several research efforts have explored AI-based monkeypox diagnosis using image classification techniques. Studies leveraging CNNs, ResNet, and EfficientNet have shown promising results in distinguishing monkeypox lesions from other dermatological conditions[8][9]. However, image-based models require high-quality datasets and suffer from misclassification in cases of similar skin conditions. The real-world deployment of image-based classifiers is often limited due to the computational cost associated with training deep learning models on large image datasets.
In contrast, symptom-based models have gained attention for their ability to detect diseases without requiring medical imaging. Previous studies have implemented gradient boosting algorithms (XGBoost, CatBoost, LightGBM) to classify diseases based on symptoms[10][11]. While boosting techniques have improved performance, they often lack explainability, limiting their real-world deployment. Moreover, traditional machine learning approaches struggle with class imbalances and feature dependencies, which can lead to biased predictions in real-world clinical applications. Some existing studies have attempted to develop symptom-based models, but their effectiveness has been constrained due to small dataset sizes, limiting their ability to generalize across diverse patient populations [12]. These limitations emphasize the need for models trained on large-scale symptom datasets, enabling improved accuracy and robustness in monkeypox diagnosis.

3. Methodology

The study follows a structured approach to develop and evaluate the Optimized Deep Neural Framework (ODNF) for monkeypox diagnosis based on clinical symptoms. This section details the dataset, preprocessing techniques, machine learning models used for comparison, and the architecture of the proposed deep learning model. The overall pipeline, showcasing how input symptom-based data is processed through various models (Random Forest, XGBoost, CatBoost, and ODNF) before interpretability techniques like LIME are applied, is depicted in Figure 2. This flowchart provides a structured view of the analytical framework employed in this study**.**

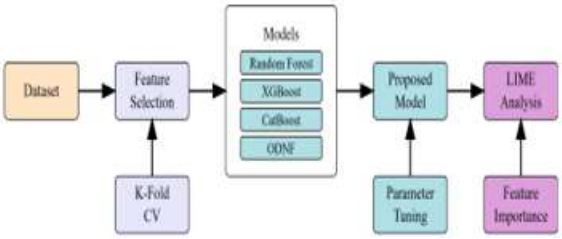

Figure 1

Dataset

The dataset utilised in this research was obtained from a Kaggle source that is openly accessible, which is based on a study published by BMJ on clinical presentations of monkeypox cases during the 2022 outbreak[13]. The original dataset contained thousands of patient records, with each entry documenting various clinical symptoms and a target variable indicating whether the patient was diagnosed with monkeypox. However, to ensure high-quality model training, the dataset underwent preprocessing to address missing values and class imbalance.
One of the primary issues was the presence of missing values in the Systemic Illness column. Rather than imputing these values, the rows containing missing entries were removed to maintain data

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, March : 2025

integrity. Additionally, the dataset was highly imbalanced, with significantly more negative cases than positive cases. This imbalance was addressed through data balancing techniques, ensuring an equal distribution of both classes to prevent biased model training. A snapshot of the dataset structure is shown in Figure 2, providing insight into the clinical symptoms used for model training.

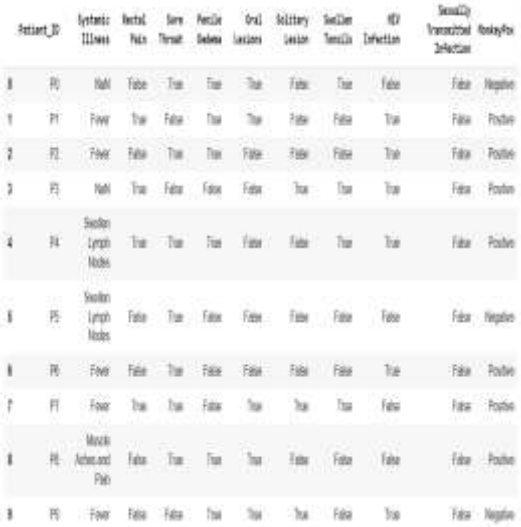

Figure 2

Machine Learning Models

A variety of machine learning models, including Random Forest, XGBoost, and CatBoost, were used to establish a performance benchmark. These models were chosen due to their proven effectiveness in medical diagnosis and their successful performance in handling structured datasets. To increase classification accuracy, Random Forest, an ensemble learning technique, constructs many decision trees and aggregates their output [14]. XGBoost, a popular gradient boosting algorithm, enhances prediction precision by iteratively rectifying mistakes from previous cycles[15]. CatBoost, a different gradient boosting technique, excels at efficiently managing categorical data and reducing overfitting[16].

Every machine learning model experienced hyperparameter adjustments to enhance its performance. A cross-validation method was utilized to guarantee reliability in assessment, reducing overfitting while enhancing generalization. The models were evaluated with standard classification metrics such as, precision, recall, accuracy F1-score, and ROC-AUC. These metrics offered a thorough assessment of the performance of conventional classifiers prior to contrasting them with the suggested deep learning model.

Proposed Model (ODNF)

The Optimized Deep Neural Framework (ODNF) was developed as a deep learning-based classification model capable of effectively predicting monkeypox cases based on clinical symptoms. In contrast to conventional machine learning algorithms that use boosting techniques or predefined decision trees, ODNF leverages deep neural networks to extract meaningful representations from patient data, improving classification accuracy. The model is structured with multiple hidden layers, allowing it to learn complex relationships between symptoms and disease outcomes. The architectural design of ODNF is depicted in Figure 3, illustrating the flow of patient data through the network. To enhance the generalization capability of ODNF, A number of optimisation strategies were used. By randomly deactivating specific neurones during training, dropout regularisation was used to lessen overfitting. In order to stabilise the learning process, batch normalisation was implemented to improve model convergence by normalizing activations at each layer. Additionally, an adaptive optimization algorithm was utilized to dynamically adjust learning rates, ensuring efficient weight updates throughout training.

To further refine model performance, early stopping was implemented, preventing

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, March : 2025

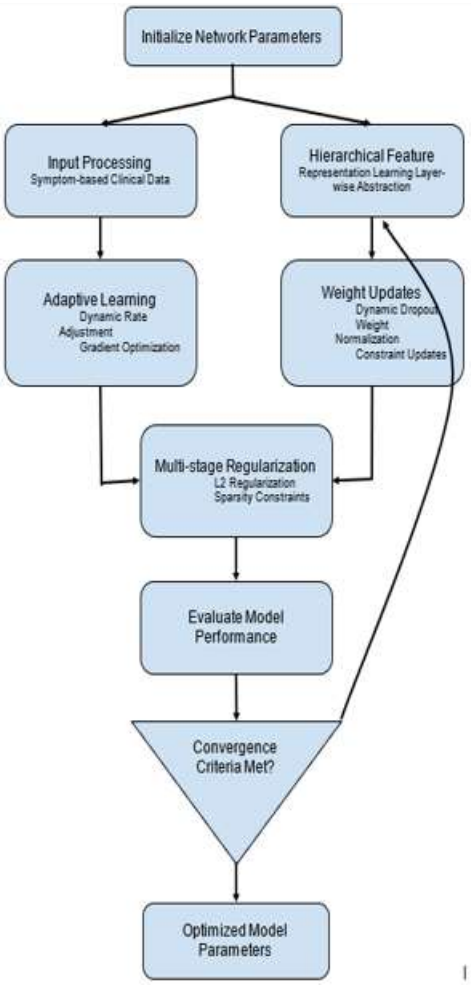

Figure 3

excessive training cycles once validation performance plateaued. A learning rate scheduler was also integrated to reduce the learning rate upon detecting stagnation in loss improvement, enabling smoother convergence. The final output layer of the model employs an activation function suitable for binary classification, allowing for accurate determination of monkeypox presence or absence.

To formally define feature transformation in ODNF, let $R^{N\times d}$ represent the input matrix, where N is the number of samples and d is the feature dimension. The transformation at the l-th layer is given by:

$H^{(l)} = f(W^{(l)} H^{(l-1)} + b^{(l)})$ (1)

Where $H^{(1-l)}$ is the activation from the previous layer, $W^l \in R^{d_l \times d_{l-1}}$ is the learned weight matrix, b^l is the bias term, f(·) represents a non-linear activation function (ReLU, Sigmoid), $H^{(0)} = X$ is the original input matrix.This formulation ensures that feature representations evolve hierarchically, capturing increasingly abstract patterns at each layer.To ensure stable weight updates, ODNF employs adaptive gradient correction, modifying the
weight update rule in gradient descent as follows:

$W_{t+1} = W_t - \eta_t \cdot (m_t \div (\sqrt{v_t} + \varepsilon))$ (2)

Where m_t and v_t are the first and second moment estimates of the gradients, ε is a tiny constant to avoid division by zero, and W_t stands for the weights at iteration t. η_t is the dynamic learning rate at iteration t.

To formally define the regularization applied in ODNF, we introduce a customized weight constraint function based on L2 regularization:

$\Omega(W) = \lambda \sum_{(i,j)} W^2_{i,j} + \alpha \sum_j |H_j|$ (3)

Where $\lambda \sum_{(i,j)} W^2_{i,j}$ penalizes large weight values (L2 regularization) and $\alpha \sum_j |H_j|$ controls activation magnitudes to prevent neuron saturation.

By integrating these multi-objective constraints, ODNF enforces a balance between representational richness and weight sparsity, preventing overfitting while ensuring optimal feature extraction.

4. Experimental Setup

The model was trained in a GPU-accelerated environment using TensorFlow and Keras, ensuring efficient computation and scalability for large datasets. To guarantee an impartial evaluation procedure, the dataset was divided into three parts: 80% for training, 10% for validation, and 10% for testing. To improve model performance while lowering the risk of overfitting, a variety of hyperparameters were carefully changed, such as batch size, learning rate, and the number of epochs. During training, a scheduling method was used to dynamically adjust the learning rate in order to obtain stable convergence.

A 10-fold cross-validation technique was used to evaluate model performance, guaranteeing that no particular data split had an impact on the results of the evaluation. In order to provide a thorough evaluation, the model's efficacy was assessed using popular classification measures, such as precision, recall, accuracy, F1-score, and the area under the ROC curve (AUC-ROC). By analysing validation accuracy, early stopping was used during training to cut down on pointless computations and avoid overfitting.

Furthermore, regularization techniques, such as dropout layers and weight decay, were incorporated to improve the model's generalization ability on unseen data. In addition to training the deep learning-based ODNF model, performance comparisons were conducted with traditional machine learning models, including Random Forest, XGBoost, and CatBoost. These machine learning models also underwent hyperparameter tuning to achieve their best possible performance.The trained model's generalization ability was evaluated using an independent dataset, and confusion matrices were used to analyze classification outcomes, particularly focusing on false positives and false negatives. These experimental evaluations provided insights into the model’s effectiveness and how well it generalizes across different data distributions.

5. Results and Discussion

A predictive model's efficacy is assessed by how well it generalises to new data while retaining high classification accuracy. This study evaluated the suggested Optimised Deep Neural Framework (ODNF) with several machine learning models, such as Random Forest, XGBoost, and CatBoost. Key classification parameters like accuracy, precision, recall, F1-score, and AUC-ROC were used to assess these models' performance.A 10-fold cross-validation method was used to guarantee generalisation and robustness. The interpretability of the suggested model, which was examined using LIME (Local Interpretable Model-Agnostic Explanations), was another important component of this investigation. The most important features in the categorisation process were found using LIME. The results in this part provide a thorough examination of feature significance and model performance.

Performance of the Proposed Model

When compared to conventional machine learning models, the suggested ODNF model performed better. With an accuracy of 99.12%, the model outperformed other models by a substantial margin. The model maintains an outstanding balance between false positives and false negatives, as seen by the high precision (98.76%) and recall (99.28%).Figure 4, which shows the accuracy and loss rates over epochs, shows the ODNF training process. The model's effective generalisation is confirmed by the Figure's fast convergence and excellent validation accuracy stabilisation.Table 1 summarises the final evaluation results on the test dataset, which show that ODNF continues to perform well in classification.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, March : 2025

Table1

Phase	Accuracy(%)	Loss
Training	99.35	0.012
Evaluation	99.92	0.018
Test Set	99.10	0.020

According to a thorough confusion matrix investigation, ODNF reduces misclassification mistakes. Monkeypox-positive patients are rarely misdiagnosed thanks to the decreased false negative rate, which is essential for containment and early management.Figure 5 is an illustration of the confusion matrix. The ODNF model's great discriminatory ability in differentiating between monkeypox-positive and negative cases is further demonstrated by its AUC-ROC score. According to the ROC-AUC curve (Figure 6), ODNF was successful in differentiating between the two groups, as evidenced by its AUC score of 0.99.

Comparative Analysis of Models

The following Table 2 presents the performance metrics of all models tested in this study

Figure 4

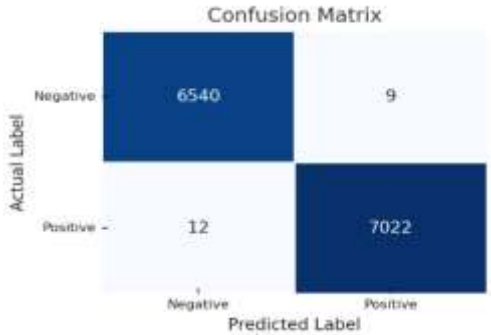

Figure 5

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 3, March : 2025

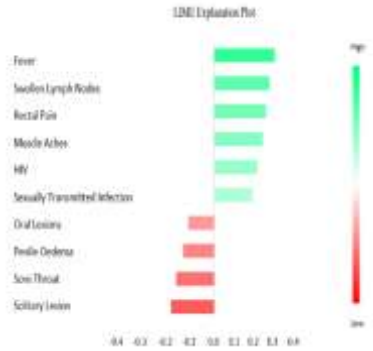

Figure 6

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
ODNF (Proposed)	99.12%	98.76%	99.28%	99.02%	0.99
XGBoost	62.00%	63.00%	62.00%	62.00%	0.61
CatBoost	64.19%	62.34%	72.36%	66.98%	0.64
Random Forest	59.29%	59.00%	59.00%	59.00%	0.58

Table 2

This comparison clearly illustrates that ODNF significantly outperforms conventional machine learning models in all key performance metrics.

K-Fold Cross-Validation

10-Fold Cross-Validation was used to assess ODNF's capacity for generalisation. The model is robust and does not overfit to the training data, as evidenced by the mean accuracy over folds being constant with little deviation. This provides more proof of ODNF's dependability in practical settings.Table 3 provides an illustration of the findings.

Fold	Accuracy	Precision	Recall	F1 Score
1	98.6%	97.9%	99.2%	98.5%
2	98.8%	98.2%	99.3%	98.7%
3	99.1%	98.7%	99.5%	99.1%
4	99.0%	98.6%	99.4%	99.0%
5	98.7%	98.1%	99.1%	98.6%
6	98.9%	98.3%	99.2%	98.8%
7	99.2%	98.9%	99.6%	99.2%
8	98.8%	98.4%	99.2%	98.8%
9	99.1%	98.8%	99.5%	99.1%
10	98.9%	98.5%	99.3%	98.9%

Table 3

6.LIME-Based Model Interpretability

The most important features influencing the classification result were determined using Local Interpretable Model-Agnostic Explanations (LIME) in order to guarantee decision-making transparency [17][18]. The results are illustrated in Figure 6, showcasing the contribution of symptoms such as fever, swollen lymph nodes, rectal pain, and muscle aches towards a positive monkeypox diagnosis.LIME analysis confirms that key clinical symptoms strongly correlate with monkeypox cases, reinforcing the model's reliability. Features such as fever, swollen lymph nodes, rectal pain, and sexually transmitted infections emerged as the most influential factors in predicting positive cases. This

interpretability ensures that medical practitioners can understand and trust the predictions made by ODNF, making it a valuable tool for real-world deployment.

7.Conclusion and Future Work

This study introduces the Optimized Deep Neural Framework (ODNF), a deep learning model for monkeypox diagnosis based on clinical symptoms. Unlike prior research relying on image-based detection or traditional boosting models, our approach enhances accessibility and interpretability. The comparative analysis demonstrated that ODNF significantly outperforms conventional models, achieving an accuracy of 99.12%. By using 10-fold cross-validation, the model's resilience was further confirmed, guaranteeing dependability in practical situations. Additionally, LIME explainability confirmed that key clinical features contributing to monkeypox diagnosis align with medical knowledge, reinforcing model transparency.

While the results highlight the efficacy of ODNF, future research can focus on integrating real-time clinical deployment for improved early detection. Expanding the dataset with multimodal inputs, such as laboratory test results, may enhance prediction accuracy. The proposed framework serves as a foundation for AI-driven infectious disease diagnosis, with potential applications in broader healthcare AI.

8.References

[1]Ladnyi, I.D.; Ziegler, P.; Kima, E.” A Human Infection Caused by Monkeypox Virus in Basankusu Territory, Democratic Republic of the Congo”. *Bull. World Health Organ.* 1972, 46, 593–597. https://iris.who.int/handle/10665/263477

[2] Zumla, A.; Valdoleiros, S.R.; Haider, N.; Asogun, D.; Ntoumi, F.; Petersen, E.; Kock, R. “Monkeypox Outbreaks Outside Endemic Regions: Scientific and Social Priorities”. *Lancet Infect. Dis.* 2022, 22, 929–931. https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(22)00354-1/fulltext

[3] Tarin-Vicente, E.J.; Alemany, A.; Agud-Dios, M.; et al.” Clinical Presentation and Virological Assessment of Confirmed Human Monkeypox Virus Cases in Spain: A Prospective Observational Cohort Study”. https://doi.org/10.1016/S0140-6736(22)01437-3

[4]Harapan, H.; Ophinni, Y.; Megawati, D.; et al. “Monkeypox: A Comprehensive Review of Transmission, Prevention, and Therapeutics”. https://doi.org/10.3390/v14102155

[5] Fink, D.L.; Callaby, H.; Luintel, A.; et al. “Clinical Features and Management of Human Monkeypox: A Systematic Review and Meta-analysis”. Lancet Infect. Dis. 2022, 309, 6–14. https://doi.org/10.1016/S1473-3099(22)00700-5

[6] Ahsan, M.; Ramiz, M.; Ali, S.; Islam, K.; Farjana, M.; Nazmus, A.; Al, K.; Akter, S. “Deep Transfer Learning Approaches for Monkeypox Disease Diagnosis”. Expert Syst. Appl. 2023, 216, 119483. https://doi.org/10.1016/j.eswa.2022.119483

[7] Bala, D.; Hossain, M.S.; et al. “MonkeyNet: A Robust Deep Convolutional Neural Network for Monkeypox Disease Detection and Classification”. Neural Netw. 2023, 161, 757–775. https://doi.org/10.1016/j.neunet.2023.02.022

[8] Jaradat, A.S.; Al Mamlook, R.E.; Almakayeel, N.; Alharbe, N.; Almuflih, A.S.; Nasayreh, A.; Gharaibeh, H.; Gharaibeh, M.; Gharaibeh, A.; Bzizi, H. “Automated Monkeypox Skin Lesion Detection Using Deep Learning and Transfer Learning Techniques”. Int. J. Environ. Res. Public Health 2023, 20, 4422. https://doi.org/10.3390/ijerph20054422

[9] Altun, M.; Gürüler, H.; Özkaraca, O.; Khan, F.; Khan, J.; Lee, Y. “Monkeypox Detection Using CNN with Transfer Learning”. Sensors 2023, 23, 1783. https://doi.org/10.3390/s23041783

[10] Kundu, D.; Siddiqi, U.R.; Rahman, M.M.” Vision Transformer Based Deep Learning Model for Monkeypox Detection”. Proceedings of the 2022 25th International Conference on Computer and Information Technology (ICCIT), Tabuk, Saudi Arabia, 25–27 January 2022. https://doi.org/10.1109/ICCIT57492.2022.10054797

[11] Malki, Z.; Atlam, E.-S.; Elhosseini, M.A.; Gad, I.; Salem, A.M. "Monkeypox Diagnosis Using Machine Learning Techniques: A Comparative Study". Comput. Biol. Med. 2023, 161, 106911.). https://doi.org/10.1016/j.compbiomed.2023.106911]
[12]Farzipour, A.; Elmi, R.; Nasiri, H."Detection of Monkeypox Cases Based on Symptoms Using XGBoost and Shapley Additive Explanations Methods". *Diagnostics* **2023**, *13*(14), 2391. https://doi.org/10.3390/diagnostics13142391
[13] Ahmed, M. Monkeypox Patients Dataset. *Kaggle*. 2022. Available online: https://www.kaggle.com/datasets/muhammad4hmed/monkeypox-patients-dataset (accessed on -25 January , 2025).
[14] Breiman, L. Random Forests. Machine Learning 2001,https://doi.org/10.1023/A:1010933404324
[15] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*. https://arxiv.org/abs/1706.06699
[16] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). "CatBoost: Unbiased Boosting with Categorical Features." *Advances in Neural Information Processing Systems (NeurIPS)*. https://arxiv.org/abs/1810.11363
[17] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier.Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*"*. https://arxiv.org/abs/1602.04938
[18] Molnar, C. (2022). Interpretable MachineLearning.https://christophm.github.io/interpretable-ml-book/