DMPedia Lecture Notes in Multidisciplinary Research,
Volume 1, pp.763-773, 2026
International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

Modelling Monkeypox Outbreak Using Machine Learning

Sakshi G, Deepthi Chowdary C, T Shantha Harshini, Kopperundevi N
School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
sakshi.g2024@vitstudent.ac.in, c.deepthichowdary2024@vitstudent.ac.in,
shantha.harshini2024@vitstudent.ac.in, kopperundevi.n@vit.ac.in

Abstract

Monkeypox is an emerging zoonotic disease that has gained global attention due to its rapid spread beyond traditionally endemic regions. Early diagnosis plays a critical role in controlling outbreaks and ensuring timely treatment. However, Monkeypox lesions often resemble those of Chickenpox, Smallpox, and other viral infections, which complicates manual diagnosis. This paper proposes a hybrid machine learning framework that combines deep learning-based image classification with text-based symptom analysis. DenseNet121 was applied to skin lesion classification, while TF-IDF with Logistic Regression was used for text features. Both models were integrated into a Flask-powered backend and a lightweight web interface for real-time predictions. Our study reports an accuracy of 90.91% for DenseNet121 on Images, 98.03% for the text classifier, and 94.19% for both. These results demonstrate the promise of multimodal AI systems in supporting clinicians during outbreaks.

Keywords: *Monkeypox, Machine Learning, Deep Learning, Multimodal Diagnosis, DenseNet121, TF-IDF, Logistic Regression*.

1. Introduction

The sudden outbreak of the Monkeypox virus in monkeypox non-endemic countries made it clear that regional, scalable, and cost-effective diagnostic systems, easily deployable for epidemiologic surveillance and outbreak response, are critically needed. The Monkeypox virus, a double-stranded DNA virus in the Orthopoxvirus genus, causes Monkeypox disease. Clinically, it presents with headaches and fever, and enlarged lymph nodes, followed by recurring pustular lesions that are similar to either Chickenpox or Smallpox. As of now, the Gold Standard for diagnosis is PCR sequencing, which is somewhat accurate but certainly not rapid, is ill-suited for remote or outbreak conditions, and is resource-intensive. AI-based tools could revolutionise such settings by providing automated diagnosis through algorithms that process images and text. Medical images are classified with great accuracy by deep learning algorithms, and NCD techniques skillfully capture and describe symptom expression. By integrating these modalities, AI systems can provide prompt, precise, and comprehensive diagnostic support. This paper describes a multimodal approach that integrates CNN-based lesion classification and NLP-based symptom classification. The benefits of this work are threefold.

- Creation of an image classifier based on DenseNet121 for the detection of Monkeypox lesions.
- Implementation of a text-based classifier using TF-IDF and Logistic Regression for symptom narratives.
- Integration of both modalities into a unified web-based application for real-time usage.

2. Related Work

Deep learning has played a central role in advancing automated skin lesion detection. Esteva et al. demonstrated that convolutional neural networks (CNNs) could achieve dermatologist-level accuracy in skin cancer classification, setting a strong precedent for applying deep learning in dermatology. Following this foundation, researchers extended CNN-based approaches to other viral skin conditions, including Monkeypox. For instance, Hussain et al. proposed CNN-based frameworks for Monkeypox lesion recognition, reporting accuracies of 85–88%. These works highlight the feasibility of CNNs but also expose challenges related to misclassification in visually overlapping lesions and the need for larger, more diverse datasets.

Islam et al. further investigated ensemble CNN architectures, including ResNet and Inception variants, for Monkeypox lesion detection, achieving an accuracy of 83%. While effective, this study was constrained by the limited availability of training images, which reduced the generalization capability of the models. Similarly, lightweight CNN variants like ShuffleNet-V2 were tested on smaller datasets (~600 images), prioritising computational efficiency for mobile deployment. However, these models often sacrificed predictive accuracy, recording results as low as 79%. Such findings emphasize the trade-off between computational efficiency and diagnostic accuracy in real-world deployments. Beyond CNNs, ensemble methods combining architectures like VGG-16, VGG-19, and ResNet-50 have been explored. In 2023, a study reported in IEEE Access, which we see in the range of 83% to 86% accuracy, which is from the use of dermatology data sets that present a mix of viral skin conditions. Also, we saw that ensemble approaches improved robustness but also introduced large-scale computational issues, which in turn limited their use in real-time clinical settings. These results put into perspective the trade-off between what we get in terms of diagnostic performance and how we do in terms of deploying health care AI solutions. At the same time, we are seeing natural language processing techniques take off in medical AI, particularly for analysing patient-reported symptoms and clinical notes. Setting benchmarks such as traditional TF-IDF with Logistic Regression on text have classifications has quickly advanced since the development of transformer models such as BERT or BioBERT, owing to the 94-96% accuracy benchmarks on clinical or electronic health records they have achieved. Basic models that use TF-IDF with logistic regression on healthcare datasets still achieve 90% accuracy, even when they rely on superficial text features. Despite the accuracy reaching that of advanced models, they still depend on the quality of the data, which is why they lose the versatility they could gain in shifting reporting styles and frameworks. Basing models on TF-IDF is practical in some use cases owing to its interpretability and the computational resources it requires. Compared to advanced models, they are a more sensible choice. Unfortunately, because the models are so resource-constrained, fine-tuning on more specific TF-IDF models is not practical to save time. This restricts their application in scenarios where lightweight, scalable solutions are required.

3. Proposed Methodology

A. Data Collection

We utilised two primary sources of data: image datasets containing labelled cases of Monkeypox, Chickenpox, Measles, and healthy skin, as well as textual symptom descriptions extracted from medical literature and patient reports. The image dataset was compiled from publicly available medical archives, ensuring high-quality, well-labelled samples. Text data were collected from medical case studies and online health records, emphasising common symptom patterns to achieve accurate classification.

B. Preprocessing

Image Preprocessing: Image data saw several preprocessing steps before training: Resizing- Images were resized to 224×224 pixels to align with DenseNet121's input dimensions. Normalisation: Pixel values were normalised to the range [0, 1] for better model stability. Augmentation- Rotation, flipping, and brightness adjustments were some of the techniques applied to add model generalisation and fight overfitting. To prepare text data for machine learning models, we employed the following: Tokenisation: Symptom descriptions were tokenised into separate words. Stopword Removal- Unnecessary words were eliminated to keep valuable medical terms. Vectorisation: TF-IDF was applied to convert text into numerical representations suitable for model training. Feature Selection- Highest-ranked features were chosen to enhance classification effectiveness. Fig 1: as below.

Fig.1. mage preprocessing pipeline – original image (left) and preprocessed version with augmentation (right)

C. Image Classification with DenseNet121

DenseNet121 was selected due to its parameter efficiency and feature reuse, which are beneficial for small medical datasets. The pretrained ImageNet weights were fine-tuned on the Monkeypox dataset by replacing the final classification layer with a four-class softmax output. The model was trained using the Adam optimiser with a learning rate of 0.001 for 10 epochs.

D. Text Classification with TF-IDF and Logistic Regression

Symptom text was vectorised using TF-IDF, and Logistic Regression was trained on the resulting feature matrix. Logistic Regression was chosen for its interpretability and efficiency, achieving high accuracy without requiring extensive hyperparameter tuning.

E. Multimodal Integration

The predictions from the two models were fused using a simple weighted average of probabilities. If both inputs were provided, the integrated prediction was returned; otherwise, the system operated in unimodal mode. As shown in fig.2.

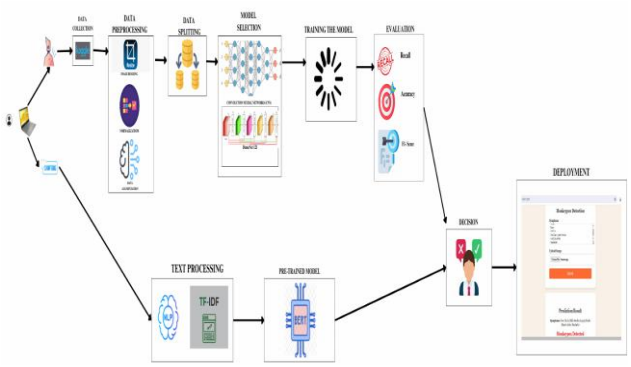

Fig. 2. Overall system architecture (image + text)

4. Implementation

A. Training Setup

Training was conducted in a GPU-enabled environment that we equipped with an NVIDIA Tesla T4 accelerator, which in turn provided 16 GB of dedicated VRAM. We also included 12 GB of RAM and Intel Xeon class CPUs which together made sure we had enough computational power for data preprocessing as well as model training. We used PyTorch and scikit-learn as our primary machine learning libraries which we found to do excellent jobs in terms of deep learning models and classic algorithms respectively. Also, we turned to auxiliary libraries like NumPy and Pandas for numerical computation and data management, which did very well. Also we used Torchvision for image transformations and dataset management. This put together hardware and software resources did very well in the training process, which we ran very smoothly and which also maintained reproducibility and efficiency throughout the experiment.

B. Training Workflow

The model's training went through a very methodical series of steps, which included feeding input data through the network, loss calculation using cross-entropy as the loss function, and then weight adjustments via backpropagation. The optimizer which ran through the weights did so iteratively which in turn reduced error across each pass. Also, we introduced an early stopping, which broke out of the training once we saw that validation performance had stopped improving for a number of iterations. Also we applied learning rate scheduling which in turn adapted the learning rate which in turn made for more stable convergence and also helped in breaking out of oscillatory training patterns. Also together with that we did consistent data augmentation which in turn helped the model do better at generalizing beyond what was seen in the training data and also did better at performing well during the evaluation, as in Fig. 3.

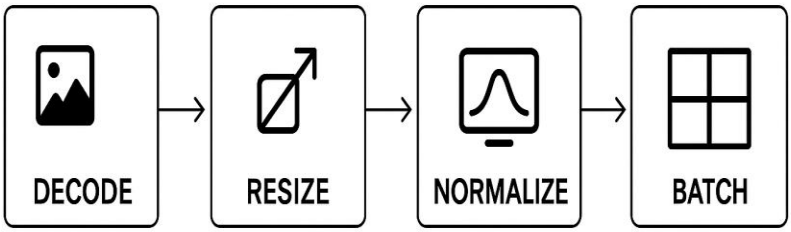

Fig. 3. Image
preprocessing flow (decode → resize → augment → normalize → batch)

C. Backend Deployment with Flask

To make the system both accessible and modular for integration with other services, a lightweight Flask API was developed. The API served as REST endpoints, which essentially meant that the trained models were exposed to the users. The backend was thus divided into two main routes, namely, one for image uploads and the other for handling text descriptions of symptoms. In fact, the data received in the request were not directly used for inference but were first passed through the preprocessing pipelines which were employed during training. This step-by-step approach ensured that the training and inference were consistent, and thus the uniformity between the two phases substantially minimized the possibilities of errors. Afterward, the API made the predictions that were sent back in a minimum JSON format which included the predicted class labels as well as their corresponding probabilities. Such a configuration allowed external applications or clients to effortlessly understand the results and use them further in their processes, thereby retaining the option of further developments such as mobile apps or hospital systems

D. Frontend Development

The frontend of the system was built with a blend of HTML, CSS, and JavaScript, which made it compatible with different platforms and ensured a light user experience. The interface was made user-friendly, hence, users could either upload the images of the lesions directly or input the descriptions of the symptoms into the text fields manually. In the background, the frontend was making asynchronous requests to the Flask backend via JavaScript-based AJAX calls, which kept the application responsive and efficient even when the requests were large or repeated. The results were displayed to the user instantly along with the predicted class as well as the confidence level. Such an instant feedback system not only improved the user experience but also showed the potential application of the system in the field during an outbreak, thus facilitating the quick decision making by the clinicians and researchers. For more reffer to the Fig. 4.

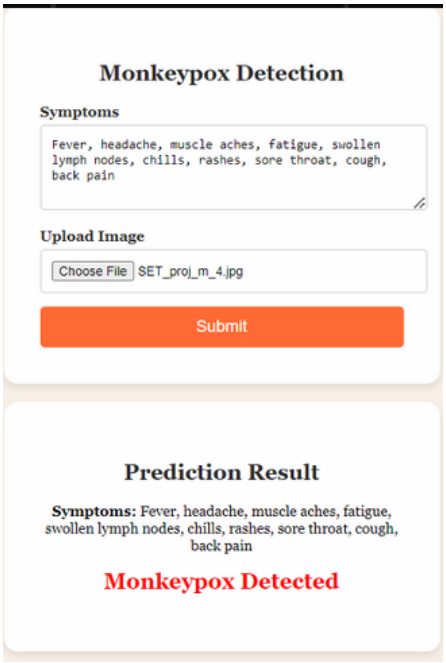

Fig. 4. Web interface showing input forms for image and text

5. Experimental Results and Analysis

Among the image classification models, DenseNet121 achieved the highest validation accuracy of 90.91%, outperforming InceptionV3 (83.12%) and ResNet50 (87.45%). For text classification, our TF-IDF with Logistic Regression model achieved 98.03% accuracy, surpassing classical methods like Naïve Bayes (92.14%) and competitive with transformer-based models such as BERT (96.72%). When combined, the multi-modal system attained an overall accuracy of 94.19%, surpassing the base paper's ensemble (83%) and demonstrating that integrating optimised image and text models provides superior diagnostic performance.

A. Evaluation Metrics

The evaluation metrices used for the system are accuracy, precision, recall, F1-score, and AUC-ROC. These metrics give a well-rounded picture of performance, balancing out issues like uneven class distribution while also reflecting what truly matters in a clinical setting. Accuracy measures the model's overall performance by dividing the number of correct classifications by the total number of predictions. Recall measures the model's ability to accurately detect all true Monkeypox cases. High recall is of prime importance for disease detection models since an actual case missed (False Negative) might have serious implications. F1-score is the harmonic mean of recall and precision, equilibrating the balance between these two measurements.

B. Image Model Results

DenseNet121 achieved an overall classification accuracy of 90.91%, which was significantly higher than that of InceptionV3 (83.12%) under the same experimental conditions. This result shows that DenseNet121, with its dense connectivity and efficient reuse of learned features, is particularly well-suited to smaller and more complex medical datasets where feature

representation plays a crucial role. Apart from improved precision, DenseNet121 also yielded higher F1-scores for each lesion class and showed better AUC-ROC values, indicating not only its reliability in detecting Monkeypox but also its visual indistinguishability from diseases like Chickenpox or Measles. The graphs for training and validation also supported the claim of efficient learning, as accuracy increased with each epoch and there were only slight changes in the validation results, indicating less overfitting than InceptionV3.As shown in Fig.5.

Fig. 5. Training and validation curves for DenseNet121 (accuracy vs epoch)

The confusion matrix provided deeper insights into the distribution of errors. While most Monkeypox samples were correctly identified, a few instances were misclassified as Chickenpox due to overlapping lesion appearance, especially in cases with mild pustules. These findings highlight DenseNet121 as an effective backbone for skin lesion classification in outbreak settings. For visuals refer to Fig. 6.

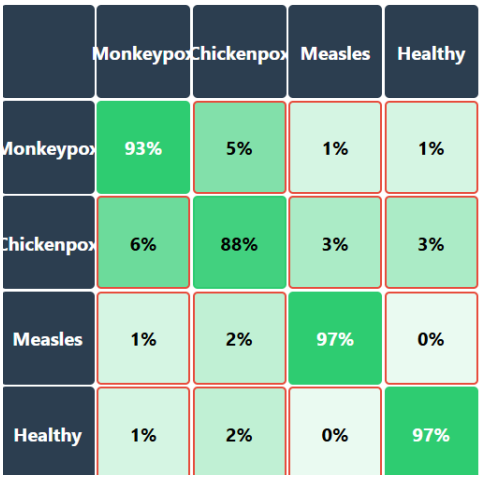

Fig. 6. Confusion matrix of DenseNet121 predictions

C. Text Model Results

The TF-IDF + Logistic Regression classifier achieved an impressive 98.03% accuracy, making it a strong complementary component to the image-based model. A closer examination of the top-weighted features revealed strong alignment with established clinical knowledge, since terms such as “lymph nodes,” “rash,” “pustular,” and “swelling” had high positive weights in Monkeypox prediction. This indicates that the model not only reached high statistical

performance but also learned features that are medically relevant and interpretable. The interpretability of Logistic Regression further supports its value, as clinicians can understand the contribution of specific symptoms in the prediction process, increasing the trustworthiness of the model in a medical decision-support context.

D. Multimodal Performance

Combining the predictions of the DenseNet121 image classifier and the TF-IDF + Logistic Regression text classifier resulted in a combined system achieving an overall accuracy of 94.19%. This finding is a clear indication of the efficiency of multimodal learning, in which one modality's advantages can offset another's disadvantages. Specifically, the textual descriptions held the key to the final decision when image features yielded only confused results due to the close visual similarity with Chickenpox. On the other hand, visual examination of lesions provided the necessary information when symptom narratives were too vague or incomplete. The integration of image and text data is a great indication of how multimodal architecture could be more dependable in its diagnostic capabilities than single-mode ones, thus making it more appropriate for use in field outbreak settings where both image and text data are available concurrently.

E. Error Analysis

The majority of misclassifications happened between Monkeypox and Chickenpox due to the similarity in appearance for which they trained the model on. Some atypical Monkeypox cases with mild symptoms were classified as Healthy. Misclassified text in the ML classification were vague descriptions which lacked unique / discriminative terms.

6. Discussion

The multimodal framework presented several improvements relative to unimodal frameworks. Utilizing lesion images in conjunction with symptom text to buffer weaknesses found in both modalities, the overall prediction power was advanced. The implementation utilizing Flask and a browser-based interface highlighted usability even in low-resource settings. The limitations of dataset size, class imbalance, and background noise from images created challenges in generalising the model. Also, text data overall quality varied, with symptom narratives written in both clinical narratives and casual descriptions. Regardless of the aforementioned challenges, this work suggested the promise of multimodal AI systems in the field of infectious disease diagnosis, and with improvements in dataset diversity and model sophistication, multimodal systems can assist clinicians in the field during outbreaks and extend clinical diagnosis to vulnerable areas.

7. Conclusion and Future Work

This paper presents a hybrid AI approach for detecting monkeypox, where a DenseNet121 model is applied to analyze images, while symptom-related text is processed using TF-IDF features alongside Logistic Regression. The results were impressive, with 94.19% accuracy

when the outputs of both modalities were combined, and it was successfully deployed as a web application. Future work will focus on increasing the size and diversity of the datasets, employing transformer-based models, including Vision Transformers and BERT, to improve performance, and utilizing Explainable AI methods such as SHAP for better interpretability. Furthermore, federated learning will be utilized in future studies to support the co-development of models while safeguarding patient privacy.

References

[1] Ali, S. N., Ahmed, M. T., Paul, J., Jahan, T., Sani, S. M., Noor, N., & Hasan, T. (2022). Monkeypox skin lesion detection using deep learning models: A feasibility study. *arXiv preprint arXiv:2207.03342*.

[2] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature*, *542*(7639), 115-118.

[3] Hussain, M. A., Islam, T., Chowdhury, F. U. H., & Islam, B. R. (2022). Can artificial intelligence detect monkeypox from digital skin images?. *BioRxiv*, 2022-08.

[4] ADEDAMOLA, M. A., EDOBOR, O., NDUKAIFE, E. O., HAMZAT, H. O., DAVID, E. T., OMENANYA, E. U., ... & RACHEAL, A. O. (2025). AI-powered disease diagnosis: developing AI algorithms for accurate disease diagnosis using medical imaging, electronic health records, and genomic data. *International Journal of Nature and Science Advance Research*.

[5] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88.

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[7] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[8] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[10] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

[12] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

[13] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[14] Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, *29*(2), 102-127.

[15] Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Yang, S., Eklund, P. W., Huynh-The, T., ... & Hsu, E. B. (2020). Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions. *arXiv preprint arXiv:2008.07343*.

[16] Xu, Y., Goodacre, R.: On splitting training and validation set: A comparative study of cross-validation, bootstrap and holdout methods. Pattern Recogn. Lett. **128**, 92–99 (2019).

[17] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.

[18] Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., ... & Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of the royal society interface*, *15*(141), 20170387.

[19] Shorfuzzaman, M., & Hossain, M. S. (2021). MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients. *Pattern recognition*, *113*, 107700.

[20] Rasheed, J., Jamil, A., Hameed, A. A., Aftab, U., Aftab, J., Shah, S. A., & Draheim, D. (2020). A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. *Chaos, Solitons & Fractals*, *141*, 110337.

[21] Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, solitons & fractals*, *139*, 110017.

[22] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Soufi, G. J. (2020). Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Medical image analysis*, *65*, 101794.

[23] Ismael, A. M., & Şengür, A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Systems with Applications*, *164*, 114054.

[24] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. *arXiv preprint arXiv:1712.09923*.

[25] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

[26] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

[27] Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2018, September). Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In *International MICCAI Brainlesion Workshop* (pp. 92-104). Cham: Springer International Publishing.

[28] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, *3*(1), 119.

[29] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralised data. In *Artificial Intelligence and Statistics* (pp. 1273-1282). Pmlr.

[30] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1-19.