Proceedings of the 6th International Conference on Inventive Research in Computing Applications (ICIRCA-2025)
DVD Part Number: CFP25N67-DVD; ISBN: 979-8-3315-2141-7

Multiple Lung Disease Classification Using Fine Tuned Transfer Learning: An Explainable AI Approach

Jyothilakshmi KN
Research Scholar ,
Department of Computer Science and Information Technology
Vels Institute of Science , Technology and Advanced Studies,
Pallavaram, Chennai, India
jyothilakshmikn782@gmail.com

R. Parameswari
Professor,
Department of Computer Science and Information Technology
Vels Institute of Science, Technology and Advanced Studies,
Pallavaram, Chennai, India
dr.r.parameswari16@gmail.com

***Abstract*—Multiple Lung Disease Classification is an important problem for health care stakeholders. Tuberculosis, COVID-19, and Pneumonia are found to be severe lung diseases. All these diseases commonly have many symptoms like cough, fever, fatigue, and other breathing difficulties. Hence, a healthcare practitioner finds it very challenging to identify which lung disease the patient belongs to. Hence, it is important to develop a Machine Learning (ML) for classifying an X Ray image into 3 categories namely Tuberculosis, COVID 19, and Pneumonia. So, we developed a ML model consists of transfer learning using pre-trained Residual Networks (Res Net) classifier, reported with a Balanced Classification Accuracy (BCA) of 98.2%, precision and recall of 98.3%, F1 Score of 0.98. We compared the proposed method with that of other transfer learning methods such as InceptionV3, DenseNet 201, and LeNet using BCA and F1 Score. Then, we applied Local Interpretable Model-Agnostic Explanations (LIME) based explainability to the developed model for gaining a better understanding of the developed model.**

Keywords—Transfer Learning, Machine Learning, LIME, InceptionV3, DenseNet 201, LeNet.

I. INTRODUCTION

Accurate classification of lung diseases such as Pneumonia, Tuberculosis (TB), and COVID-19 is essential due to their significant public health impact and overlapping symptoms. These diseases often present with similar respiratory symptoms, including cough, fever, fatigue, and shortness of breath, making clinical diagnosis based solely on symptoms and physical examination extremely challenging [1]. Misdiagnosis can delay treatment and potentially lead to complications, increased transmission (in the case of infectious diseases like TB and COVID-19), and unnecessary burden on healthcare systems.

Differentiating these diseases early is also vital for infection control and clinical management. COVID-19 and TB are contagious, while pneumonia (particularly bacterial pneumonia) can be life-threatening if not treated promptly. Inappropriate classification could result in incorrect isolation protocols or antimicrobial misuse, increasing the risk of antibiotic resistance [2]. Moreover, some patients may be co-infected, requiring nuanced diagnosis and therapeutic planning that relies on accurate classification strategies.

From a health system perspective, proper disease identification allows for better triaging of patients, prioritizing resource allocation like hospital beds, oxygen supplies, and ventilators. This is particularly important during epidemics or seasonal spikes when healthcare facilities are under pressure. A well-defined classification model for these three diseases can assist in real-time surveillance, tracking disease trends, and informing public health policies [3].

Machine Learning (ML) offers a transformative solution for classifying Pneumonia, Tuberculosis, and COVID-19 by enabling automated, rapid, and accurate analysis of chest X-rays, CT scans, and other diagnostic data. Unlike traditional diagnostic approaches, ML models can learn intricate patterns and subtle distinctions in imaging features that might be missed by human experts, especially under fatigue or in high-volume clinical settings [4]. These capabilities make ML especially valuable in settings where experienced radiologists are not readily available.

Additionally, ML-based models support real-time diagnosis, which is critical during health emergencies like the COVID-19 pandemic. These models can process thousands of images per day with high throughput, facilitating early screening and triage. When deployed in low-resource environments, mobile ML applications can assist frontline healthcare workers with decision support tools that recommend likely diagnoses and confidence levels [5]. This augments clinical care and enhances diagnostic consistency across different regions and providers.

The use of interpretable ML methods like LIME or SHAP ensures that these models are not just accurate, but also transparent and trustworthy, addressing a key concern in healthcare AI applications. These techniques can highlight which regions in a lung scan contributed most to the model's decision, providing explainability to clinicians and increasing their willingness to integrate ML systems into clinical practice [6]. With continued validation and responsible deployment, ML models hold the potential to democratize expert-level diagnostics globally.

II. RELATED WORKS

This section contains the information about the previous studies conducted on the detection of multiple lung diseases Classification. We observed that many deep learning studies are used for the early detection of lung diseases such as Tuberculosis (TB), COVID-19, and Bacterial Pneumonia. Researchers utilized Convolutional Neural Networks (CNN) to classify TB, COVID-19, bacterial pneumonia. We classified the existing literature spanning from 2023 onwards into mainly 4 categories. They are:

- Standard CNN, where the researchers used the standard CNN architecture for distinguishing Tuberculosis, COVID-19, Bacterial Pneumonia, Viral Pneumonia, and Healthy Control patients,

- Pre-trained CNN architectures where the researchers utilized an existing implemented CNN network for developing an architecture, It is also known as transfer learning from the literature. For example, researchers utilized an existing pre-trained model that works pretty well on some other similar disease datasets can be efficiently utilized for the given underlying problem after minor finetuning of the model,
- The attention mechanism refers to a technique where researchers utilize specially designed CNN networks that focus more on certain regions of an input image, assigning greater importance and weight to those specific areas.

A. Conventional CNN Architectures for Lung Disease Classification

M. Hog et al. introduced a CNN model incorporating deep max pooling and softmax layers to classify lung conditions into two categories: healthy and diseased. Their model achieved a testing accuracy of 87% [7]. Another study [8] implemented an optimized neural network combined with a gradient descent algorithm to identify lung diseases such as pneumonia, lung cancer, and tuberculosis, reaching an accuracy of 99.2%. In [9], researchers applied an enhanced CNN integrated with an extreme learning machine, attaining 96.2% accuracy in classifying different stages of lung cancer severity. While these methods emphasize the architectural design of neural networks, they lack detailed explanations and modeling approaches that target specific regions within the input images.

B. Pre-trained CNN Architectures – Transfer learning

In pre-trained CNN architectures, model development leverages existing models that have already been trained on relevant datasets for classification tasks. Specifically, a pre-trained model developed using lung cancer datasets was adapted to address the current problem, achieving an accuracy of 93% in distinguishing between carcinoma and non-carcinoma cases [10]. In another study [11], researchers applied a transfer learning approach for lung carcinoma detection. This involved using pre-established VGG16 and VGG19 networks that had already been trained on large-scale datasets. Another research by Michael et al. [12] employed a deep residual network architecture to identify pneumonia and tuberculosis from chest X-ray images, achieving an accuracy of 86.4% on the dataset. Mahmud et al. [13] proposed an interpretable deep learning model aimed at the early detection of lung diseases using chest X-rays, reporting an accuracy of 91%. Although this model incorporated interpretability, it did not provide insights into the specific regions of the chest X-ray that contribute to the diagnosis of lung diseases. This limitation is consistent across all the studies referenced here [10], [11], [12], [13].

A. Research Gap and Motivation for the Study:
The previous studies used standard CNN architectures for the development of the model. However, there is a huge need for the development of efficient models that can operate on already existing pre-trained neural network models. Using pre-trained neural network models offers several advantages, particularly in domains like computer vision and medical imaging. These models significantly reduce training time and computational costs since they have already been trained on large, diverse datasets such as ImageNet. This allows researchers to fine-tune them for specific tasks without starting from scratch, making them ideal for projects with limited data or resources. Pre-trained models also provide strong baseline performance and good generalization capabilities, as their early layers capture fundamental features like edges and textures that are transferable across domains [14]. Additionally, they enable efficient experimentation with proven architectures like ResNet, and support transfer learning by applying knowledge from one domain to another, which is especially useful in cases with small, specialized datasets [14]. Overall, leveraging pre-trained models accelerates development, improves performance, and simplifies the process of building robust deep learning solutions. Thus, these solutions can act as a free capsule for developing new models [14].

III. METHODOLOGY

The overall process involves the following methodology:

- Data acquisition: The study is planned around a dataset named “Lungs Disease Dataset”, which is available in Kaggle,
- We performed augmentation of the training dataset where in 80% of the data belongs to training set and rest of the 20% belongs to the testing set,
- The overall image is resized to a size of 620*620 pixels in size,
- Then, we applied a transfer learning technique on top of the resulting data.

Table 1 illustrates the total count of the labels in the dataset.

Class Name	Count	Testing	Training	Validation
COVID-19	2510	503	1506	501
Tubeculosis	2509	504	1503	502
Pneumonia	2561	512	1540	509

A. Residual Network Neural Network:

Residual Neural Networks (ResNets) are a class of deep learning architectures that address the challenges of training very deep neural networks, particularly the vanishing gradient problem. Introduced by Kaiming He and colleagues in 2015, ResNets incorporate "skip connections" or "identity shortcuts" that allow the input of a layer to bypass one or more subsequent layers and be added directly to the output [15]. This design enables the network to learn residual functions—differences between the desired output and the input—rather than attempting to learn unreferenced mappings, facilitating the training of networks with significantly more layers. By mitigating issues like degradation of training accuracy in deeper networks, ResNets have achieved remarkable success in various applications, including image classification, object detection, and segmentation [15], [16]. Variants like ResNet-50, ResNet-101, and ResNet-152, which denote the number of layers, have become standard benchmarks in computer vision tasks.

A Residual Neural Network (ResNet) is a deep learning architecture that incorporates residual connections, allowing

the network to learn residual functions with reference to the layer inputs. This design facilitates the training of very deep networks by mitigating issues like the vanishing gradient problem. In a typical ResNet architecture, the network is divided into stages, each comprising multiple residual blocks [17]. Each block contains two or more convolutional layers, often with 3×3 kernels, and a skip connection that adds the input of the block to its output. Project projection shortcuts (1×1 convolutions) are used when necessary to manage changes in feature map dimensions. Activation functions like Rectified Linear Unit (ReLU) are applied after each convolutional layer, and batch normalization is commonly used to stabilize and accelerate training. The depth of ResNet architectures can vary; for instance, ResNet-34 consists of 34 layers, while ResNet-50 uses a bottleneck design with 50 layers. These architectures have been successful in various computer vision tasks, including image classification and object detection, due to their ability to train deep networks and capture complex patterns in data effectively. Table 2 illustrates the overall configuration and structure of the RES NET classifier with detailed information on the layer name, output size, configuration, and number of blocks.

Table 2 Residual Network Layers

Layer Name	Output Size	Configuration	Number of Blocks
Conv1	112*112	7*7 Convolution, 64 filters, stride 2	1
Max Pool	56*56	3*3 max pooling, stride 2	-
Conv2_x	56*56	3*3 convolution, 64 filters	3
Conv3_x	28*28	3*3 convolution, 128 filters	4
Conv4_x	14*14	3*3 convolution, 256 filters	6
Conv5_x	7*7	3*3 Convolution, 512 filters	3
Avg Pool	1*1	Global Average Pooling	-
Fully Connected	1*1	1000 dimension output, softmax activation	1

LIME (Local Interpretable Model-Agnostic Explanations) is an explainable AI technique designed to interpret the predictions of complex, black-box machine learning models. It operates by generating numerous perturbed versions of a specific input instance and observing how the model's predictions change in response to these slight modifications [18]. By analyzing these variations, LIME fits a simple, interpretable model—such as a linear regression—to approximate the behavior of the complex model in the local vicinity of the instance. This approach provides insights into which features most influence the model's prediction for that particular case, enhancing transparency and trust in AI systems [18], [19].

IV. Results and Findings

Table 3 illustrates the results after using the proposed Res Net classifier on the lungs disease dataset into three categories namely: COVID 19, Pneumonia, and Tuberculosis.

Table 3: Comparison of performance metrics

Classifier	BCA (%)	Precision (%)	Recall (%)	F1 Score (%)
Proposed Res Net	98%	98%	98%	98
InceptionV3	96%	96%	96%	96
LeNet	89%	89%	86%	88
EfficientNet-B0	97%	97%	97%	97

As illustrated in the Table 3, the proposed Res Net classifier reported with the highest BCA, Precision, Recall, and F1 Score of 98%, 98%, 98%, and 0.98 respectively on the testing unseen data. ResNet performs better due to its residual connections, which help mitigate the vanishing gradient problem and allow the model to effectively train deeper networks without performance degradation. This enables it to learn more complex and abstract features, making it well-suited for distinguishing subtle differences in medical imaging data.

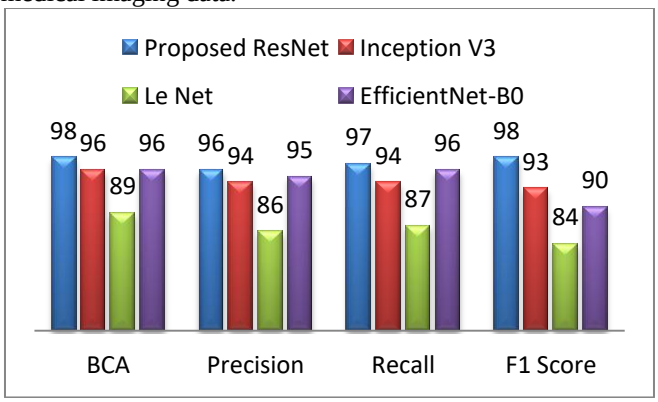

Figure 1: Comparison of Base Res Net with that of the Fine Tuned Res Net

Figure 1 has the illustration of Base ResNet with that of the Fine Tuned ResNet model. Relative to the base model, the fine tuned model is able to attain relatively higher BCA, Precision, Recall, and F1 Score. The fine-tuned ResNet model demonstrated superior performance compared to the base ResNet, achieving higher scores across all key metrics. This improvement can be attributed to several critical factors. Fine-tuning allowed the model to adapt its pretrained weights to domain-specific features present in chest X-ray images, capturing subtle variations across conditions such as COVID-19, tuberculosis, and pneumonia. By unfreezing the top residual blocks and retraining them on the target dataset, the model developed more discriminative feature representations. Additionally, the use of an optimized training head with batch normalization and dropout improved regularization and generalization. Leveraging adaptive optimizers like Adam and incorporating learning rate scheduling further ensured

efficient convergence and stability. These design choices collectively enabled the model to make more accurate and balanced predictions across all three disease classes.

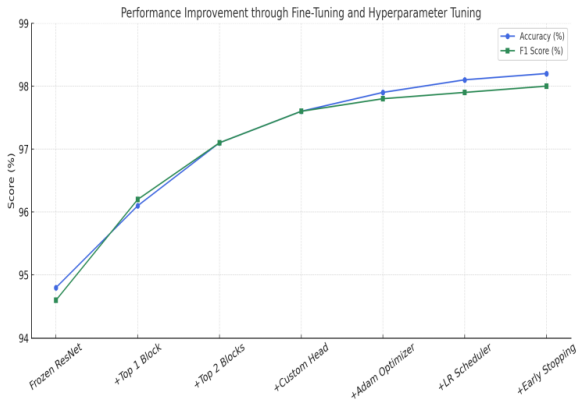

Figure 2: Hyperparameter tuning result of the ResNet.

Figure 2 has the clear illustration for the hyperparameter tuning results using the Res Net model. After hyperparameter tuning, we see a clear improvement in the accuracy, F1 Score of the base models. Fine-tuning the top 1 or 2 blocks of a Convolutional Neural Network (CNN) can improve results by leveraging the pre-learned low-level features captured in the initial layers, such as edges and textures, which are generally applicable across tasks. The higher layers, which capture more complex task-specific features, can be adjusted to better align with the new dataset, allowing the network to focus on relevant aspects of the task. This approach also reduces the risk of overfitting by updating fewer parameters, maintaining generalization while optimizing for the task at hand. Moreover, fine-tuning fewer layers lowers computational costs and accelerates convergence, as the earlier layers are already well-suited for feature extraction, thus allowing the model to quickly adapt to the specific problem without needing to retrain from scratch.

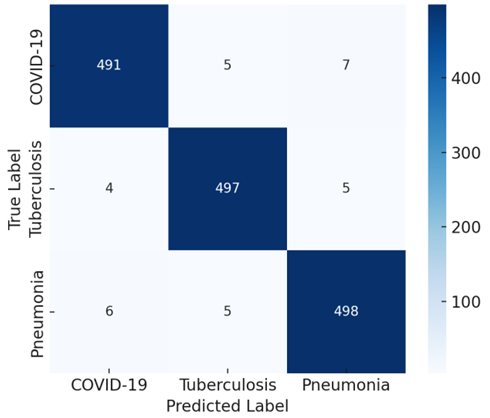

Figure 3: Illustration of CM using Res Net.

Figure 3 has the illustration of the Confusion Matrix (CM) associated with the Res Net classifier for categorizing Pneumonia, Tuberculosis, and COVID 19 patients. As illustrated in the figure 3, the highest number of misclassification is happeed when classifying pneumonia as COVID 19 and vice versa. However, there is a larger and better rate of accurate classification using the proposed model after using the Res Net architecture. There are a lot of similar symptoms that can be associated with COVID 19 and Pneumonia, which is very difficult to capture even using fine tuned Res Net classifier. The label wise improvement in the classification accuracy is a huge factor as even a slightest increase in the BCA means a lot in health care sector.

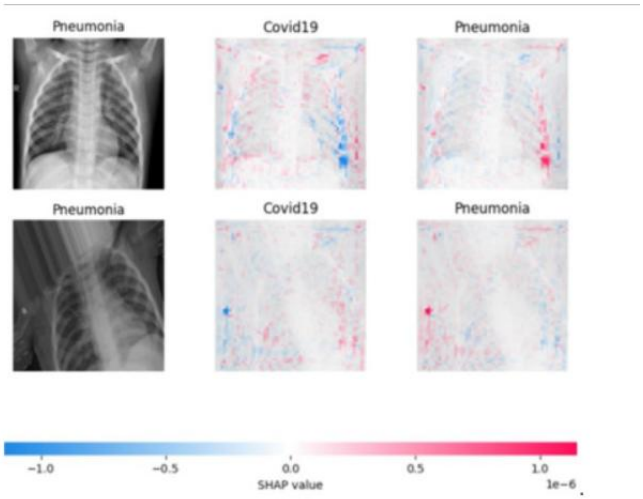

Figure 4: LIME explainer for COVID 19 and Pneumonia.

Figure 4 illustrates an explanation for the detection of COVID 19 and Pneumonia using the LIME explainer. LIME explainer can give the significant changes with respect to intensity in the form of various shades in the Chest X Ray that are crucial in distinguishing the images. The darker red shades in the images indicate a higher significant change of pixel values in those regions. Figure 4 illustrates such changes for 2 selected samples from the study.

The limitations of the study are as follows:

- The study is conducted on the Chest X Ray images with a reasonable amount of samples. However, it is important to check the model with a highly clinical image Chest X ray dataset to address the robustness and generalizability of the model.
- We classified the Chest X-rays into only 3 categories. However, there are a plethora of Chest and lung diseases where we need to apply the model to. For instance, the physicians are more interested in finding out multiple lung diseases such as: Pneumonia, Tuberculosis, COVID-19, Asthma and many more. Hence, it is a challenging task to build a classifier that can distinguish with all these multiple levels of chest diseases,
- We are also planning to consider other modalities for Chest X ray disease classification. It would be interesting to classify the Chest X Ray diseases from EEG and other sources of modalities,
- We are also planning to implement new explainability methods to understand the interpretability of the built models in a greater view.

V. CONCLUSION:

This study used a hyperparameter fine-tuned Res Net Classifier for distinguishing lungs diseases from Chest X Rays into 3 categories namely Pneumonia, Tuberculosis, and COVID 19, reported with a BCA, Precision, Recall, and

F1 Score of 98%, 98%, 98%, and 0.98 respectively. We found a key observation that the highest misclassification for the proposed method is seen when classifying COVID - 19 as Pneumonia and vice versa. Even though this is the case, there is a significant higher amount of increase in BCA and F1 score for distinguishing these three categories. The proposed model outperformed other state of the art transfer learning models such as InceptionV3, LeNet, EfficientBO.

REFERENCES

[1] World Health Organization. (2021). *Tuberculosis and COVID-19*. Retrieved from https://www.who.int/

[2] Cannesson, Alexandre, and Narcisse Elenga. "Community‐Acquired Pneumonia Requiring Hospitalization among French Guianese Children." International Journal of Pediatrics 2021, no. 1 (2021): 4358818.

[3] Flynn, JoAnne L., and John Chan. "Immune cell interactions in tuberculosis." *Cell* 185, no. 25 (2022): 4682-4702.

[4] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Ng, A. Y. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Medicine*, 15(11), e1002686. https://doi.org/10.1371/journal.pmed.1002686

[5] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 43, 635–640. https://doi.org/10.1007/s13246-020-00865-4

[6] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "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). https://doi.org/10.1145/2939672.2939778

[7] S Ashwini, J. R. Arunkumar, R Tandaihah Prabhu, N. H. Singh, N. P. Singh. "Diagnosis and Multiclassification of Lung Diseases in CXR inages using Optmized Deep Convolutionial Neural Networks", Soft Computing, Volume 28, page no: 6219-9233. 2023.

[8] S Agarwal, K. V Arya, Y. K. Meena. "CNN-0-ELMNET: Optimized Lightweight and Genralized Model for Lung Disease Classification and Severity Asessment", IEEE Transactions on Medical Imaging, Vol:43, Issue:12, 2024.

[9] M Humayun, R Sujatha, S. N. Almauayqil, N. Z. Jhanji. "A Transfer Learning Approach with Convolutional Neural Network for the Classification of Lung Carcinoma", Healthcare, 10(6), 1058, 2022.

[10] M Humayun, R Sujatha, S. N. Almauayqil, N. Z. Jhanji. "A Transfer Learning Approach with Convolutional Neural Network for the Classification of Lung Carcinoma", Healthcare, 10(6), 1058, 2022.

[11] M. Mamalakis, A. J. Swift, B. Vorselaars, S. Ray, S, Weeks, W. Ding, R. H. Clayton, L. S. Mckinzee, A. Banarjee."DensResCov19-A Deep Transfer Learning Approach for Robust Automatic Classification of COVID-19, Pneumonia, and Tuberculosis from X-rays", Volume 94, Computerized Medical Imaging and Graphics, Volume 94, 102008, December 2021.

[12] E. Mahmud, N. Fahad, M. Assadussaman, S. M. Zain, K. O.M Ghoh, M. K. Korol, An explainable Artificial Intelligence for Multiple Lung Diseases Classification from Chest X Ray Inages Using Fine Tuned Transfer Learning, Decision Analytics Journal, Volume 12, September 2024.

[13] B. Tenguio, C. Fangzhou, X. Li, "Self Attention based Speaker Recognition Using Cluster Range Loss", Neurocomputing, volume 368, pp:59-68, 2019.

[14] Salehi AW, Khan S, Gupta G, Alabduallah BI, Almjally A, Alsolai H, Siddiqui T, Mellit A. A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability. 2023 Mar 29;15(7):5930.

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

[16] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1492-1500).

[17] Bressem KK, Adams LC, Erxleben C, Hamm B, Niehues SM, Vahldiek JL. Comparing different deep learning architectures for classification of chest radiographs. Scientific reports. 2020 Aug 12;10(1):13590.

[18] Holzinger A, Saranti A, Molnar C, Biecek P, Samek W. Explainable AI methods-a brief overview. InInternational workshop on extending explainable AI beyond deep models and classifiers 2020 Jul 18 (pp. 13-38). Cham: Springer International Publishing.

[19] Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M. Explainable AI for retinoblastoma diagnosis: interpreting deep learning models with LIME and SHAP. Diagnostics. 2023 Jun 1;13(11):1932.