**Research Paper Review**

**Review 10 Papers (Latest)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Reference (Authors, papers etc.) | Type of Transformers | Database Used | Methodology Used | Limitations of Transformers | Advantages | Performance/Accuracy of Model |
| 1.  **xViTCOS: Explainable Vision Transformer Based COVID-19 Screening Using Radiography**  (ARNAB KUMAR MONDAL, ARNAB BHATTACHARJEE, PARAG SINGLA, A. P. PRATHOSH) | Vision Transformers (ViT-B/16, ViT-B/32,  ViT-L/16 and ViT-L/32) | CT Scan Dataset (COVIDx CT-2A dataset), Chest X-ray (CXR) Dataset (COVIDx CXR-2), CheXpert | Multi-stage transfer learning approach, Gradient Attention Rollout algorithm (For Explanability) | xViTCOS-CT fails to predict the ground truth  non-COVID-19 Pneumonia with confidence | * Proposed method is most balanced amongst all the baseline models. * High Specificity & NPV value is useful in clinical scenarios as it helps in efficient utilization of limited resources. * Utilized explanability- driven heatmap plot for predictive decision | xViTCOS-CT: Accuracy- 98.1%, Recall (Sensitivity)- 96%, Precision- 96%, F1-Score- 96.1%, Specificity- 98.8% & NPV- 98.8% for COVID-19 cases  xViTCOS-CXR: Accuracy- 96%, Recall (Sensitivity)- 100%, Precision- 99%, F1-Score- 99.5%, Specificity- 99.7% & NPV- 100% for COVID-19 cases |
| 2. **CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR** (Maram Mahmoud A. Monshi, Josiah Poon, Vera Chung, Fahad Mahmoud Monshi) | VGG-16, VGG-19, ResNet-18,  ResNet-34, ResNet-50, and EfficientNet-B0. | COVIDx, COVIDcxr | Data augmentation on CXR (  resizing,  flipping, rotating, zooming, warping, lighting, and normalizing) | CovidXrayNet is still at a research stage and is not suitable for direct  clinical diagnosis. | Data augmentation method & CNN hyperparameter tuning enhanced the accuracy of COVID-19 classification from CXRs | CovidXrayNet: Accuracy- 95.82%, Recall (Sensitivity)- 95.43%, Precision- 96.93%, F1-Score- 96.16%, MCC- 92.24% for COVID-19 cases |
| 3. **Explainable Vision Transformers and Radiomics for COVID-19**  **Detection in Chest X-rays** (Mohamed Chetoui and Moulay A. Akhloufi) | Vision Transformers (ViT-B16, ViT  B32,and ViT-L32) | SIIM-FISABIO-RSNA COVID-19, RSNA | Data augmentation was applied during training, Hyperparameter tuning of ViT models & visualized the signs detected by ViT by using the attention map of the best model (ViT-B32). | Model will  have an even better performance if compared to a manual interpretation by radiologists. | * Demonstrated that the obtained results outperformed comparable state-of-the-art models for detecting COVID-19 on CXR images using CNN architectures * Compared to Transformer-based models, the proposed fine-tuned ViT-B32 model showed statistically better performance than other Transformer-based models * The attention map for the Vision Transformer model showed that ViT-B32 is efficient in identifying the most important pathology regions (the signs of COVID-19) and other Pneumonia signs | ViT-B32: Accuracy- 96%, Recall (Sensitivity)- 96%, Specificity- 96% & AUC- 99% for COVID-19 cases |
| 4. **Vision Transformer based COVID-19 Detection using Chest X-rays** (Koushik Sivarama Krishnan & Karthik Sivarama Krishnan) | Vision Transformer (ViT-B/32) | COVID19 X-ray database dataset (large X-ray dataset (COVQU) & customized datasets from different sources), COVID19 Pneumonia Normal Chest X-ray PA Dataset | Transfer Learning, Image augmentation, Image enhancement methods & Hyperparameter tuning of Vision Transformers | * Noise is a key factor in radiography that affects the model’s performance * An ensemble learning-based approach can enhance the final Performance * Using a large version of ViT with a larger dataset can surely reach higher performance metrics | * The ViT model took around 10 minutes to train while other CNN based models took around 35 minutes (transformer based approaches consume less time when compared to CNN based approaches) * ViT baseline model with 32x32 as input patch size is selected as the best model for classifying COVID-19 X-ray images. | ViT-B/32: Accuracy- 97.61%, Precision- 95.34%, Recall (Sensitivity)- 93.84% & F1-Score- 94.58% for COVID-19 cases |
| 5. **COVID-Transformer: Interpretable COVID-19 Detection Using**  **Vision Transformer for Healthcare** (Debaditya Shome, T. Kar, Sachi Nandan Mohanty, Prayag Tiwari, Khan Muhammad,  Abdullah AlTameem, Yazhou Zhang and Abdul Khader Jilani Saudagar) | Vision Transformer (ViT L-16) | Constructed a three-class data set of 30 K chest X-ray pictures (Qi et al., El-Shafai et al. & Sait et al.) | Designed a COVID-19 detection pipeline utilizing the Vision Transformer  model and fine-tuned it on their dataset with a custom MLP block, Resizing & Interpolation of images, Data Augmentation techniques (random rotation, width shift, height shifts, and  flipping) | Model results in overfitting beyond 2 dense layers | * Proposed method for X-ray based detection of COVID-19 would be an efficient addition to the healthcare system boosting the global health coverage * A Grad-CAM based visualization makes their approach interpretable by radiologists and can be used to monitor the progression of the disease in the affected lungs, assisting healthcare * This may be extremely valuable in areas where quick testing is unavailable, and it may also be used as a second screening method after the standard RT-PCR test to verify that any true negative or false positive cases do not occur | COVID-Transformer: Accuracy- 92%, Precision- 93%, Recall (Sensitivity)- 89%, F1-Score- 91% & AUC- 98% for COVID-19 cases |
| 6. **COVID-19 Detection in CT/X-ray Imagery Using Vision Transformers** (Mohamad Mahmoud Al Rahhal, Yakoub Bazi, Rami M. Jomaa, Ahmad AlShibli, Naif Alajlan,  Mohamed Lamine Mekhalfi and Farid Melgani) | Vision Transformers (DeiT Architecture) | COVIDx, SARS-CoV-2 CT | Model is based on a Data-Efficient Image Transformer (DeiT)  Architecture, Token & Distiller classifiers are used, Siamese encoder is used | Limited training data was used | * Proposed system also exhibits good robustness when a small portion of training data is allocated * Generated adversarial examples, which clearly improved the performance * Proposed framework has demonstrated its robustness under limited training data * Heat maps were used which displayed the progression of focus areas over network layers, similar to X-ray images or CT images | For CXR: Accuracy- 94.62%, Precision- 96.77%, Recall (Sensitivity)- 96.77%, F1-Score- 96.77% & Specificity- 99.65% for COVID-19 cases  For CT: Accuracy- 99.13%, Precision- 99.46%, Recall (Sensitivity)- 98.82%, F1-Score- 99.13% & Specificity- 99.47% for COVID-19 cases |
| 7. **COVID-19 Detection in Chest X-ray Images Using**  **Swin-Transformer and Transformer in Transformer** (Juntao Jiang & Shuyi Lin) | Swin Transformer (Swin-B) + Transformer in Transformer (small) | Challenge of Chest XR COVID-19 detection (Dataset) | Data Augmentation (horizontal flipping, and the rotation or translation), applied label smoothing in loss & Hyperparameter tuning of model | Larger models should have better performance but failed to try them due to limitation of time | Swin Transformer and Transformer in Transformer (TNT) are successful works to adapt Transformer from language to vision and achieved the state of the art in different tasks | Results for CXR: Accuracy- 94.75%, Recall (Sensitivity)- 94.75% & Specificity- 95.09% for COVID-19 cases |
| 8**. MXT: A New Variant of Pyramid Vision Transformer for Multi‑label**  **Chest X‑ray Image Classification** (Xiaoben Jiang, Yu Zhu, Gan Cai, Bingbing Zheng, Dawei Yang) | Pyramid  vision Transformer | Chest  X-ray14 & Catheter dataset | Multi-layer overlap patch (MLOP)  Embedding, class token Transformer block and multi-label attention (MLA) are utilized, Downsampling spatial reduction attention (DSRA) & Dynamic position  feed forward with zero paddings is utilized | The heatmaps generated  from the Model-D are discrete and blurry (i.e., CXR  image with Atelectasis, Pneumothorax), and location of the  lesion regions is in error | * MXT can assist radiologists in diagnoses of lung diseases and check the placement of catheters, which can reduce the work pressure of medical staff | For Chest X-ray14: AUC- 83%, HL- 6.9% & LRAP- 81%  For Catheter: AUC- 94.6%, OE- 14.5%, HL- 5.9% & LRAP- 90.1% |
| 9. **Towards robust diagnosis**  **of COVID‑19 using vision**  **self‑attention transformer** (Fozia Mehboob, Abdul Rauf, Richard Jiang, Abdul Khader Jilani Saudagar,  Khalid Mahmood Malik, Muhammad Badruddin Khan, Mozaherul Hoque Abdul Hasnat,  Abdullah AlTameem & Mohammed AlKhathami) | Vision Transformer | Brazilian dataset (SARS‑COV2 CT scan dataset) & COVID-19 (HUST-19) | Segmentation of image is performed using transformer encoder/ decoder architecture, Layer normalization is implemented & Data augmentation (image flipping  (horizontal), resizing (image size) and rotation) | Binary dataset is heterogeneous and number of samples of CT scan images belonging to  each variation are low | * Proposed transformer vision approach can predict the quantification of COVID-19 based on the pixel values in the long-range relation-based maps * self-attention transformer-based approach is of paramount significance for the methods intent to diagnose the COVID-19 in CT scan images * This can provide the assistance to clinicians in decision making with respect to the assessment of the severity of the COVID-19 | Brazilian Data-set: Accuracy- 98%, AUROC- 99.6%  Hust19 Data-set: Accuracy- 99.7%, AUROC- 99.7% |
| 10. **Vision transformer and explainable**  **transfer learning models for auto**  **detection of kidney cyst, stone**  **and tumor from CT‑radiography**  (Md Nazmul Islam, Mehedi Hasan, Md. Kabir Hossain, Md. Golam Rabiul Alam,  Md Zia Uddin & Ahmet Soylu) | Vision Transformer (EANet, CCT, and Swin transformers) | Dataset was collected from PACS (Picture archiving and communication system)  and workstations from a hospital in Dhaka, Bangladesh | Scaled, resized, randomized images for various deep learning & transformer models. Image augmentation, normalization & Transfer learning | Resnet model in their study is the least effective at detecting kidney  tumors and kidney stones, VGG16 is watching a very small region to make a decision & Inception V3 is also not  watching the region of interest perfectly and watching more low-level features | * As compared to all the models, Swin transformer has taken less time to train with the same number of epochs * VGG16 model is better compared to Resnet50 and Inceptionv3 by showing the desired abnormalities in the anatomy better. * Superior accuracy of our model based on the Swin transformer and the VGG16-based model can both be of great use in detecting kidney tumors, cysts, and stones, and can reduce the pain and suffering of patients. | EANet:  Accuracy- 77.02%  Swin Transformers:  Accuracy- 99.30%  CCT:  Accuracy- 96.54% |