Enhanced Lung Disease Detection using VGG and Spatial Transform Network (STN)

Prasanthi A. Yavanamandha [1,*], Srikiran B. Kavuri [2], Sai Moulika C. Bedadhala [1], Vaishnavi D. Chintaparti[1], and Sindhu E. Kakumani [1]

[1] Department of CSE-AIML & IoT, VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, India
[2] Department of CICS, University of Massachusetts, Amherst, USA
Email: prasanthi_y@vnrvjiet.in (F.A.L.); srikirankavuri@gmail.com (F.B.L.); saimoulikabedadhala@gmail.com (F.C.L.); vaishnavichintaparti@gmail.com (F.D.L.); sindhukakumani20@gmail.com (F.E.L.)
*Corresponding author

***Abstract*—Pneumonia and tuberculosis are crucial reasons for infection and mortality worldwide, making lung illnesses a first-rate international fitness problem. Appropriate and active prognosis of those ailments is important for a successful remedy and for the care of the affected person. Because there are numerous similarities between lung illnesses and versions inside an unmarried condition, it may be tough to as it should be diagnosing lung illnesses. In this project, we endorse a singular technique to lung da sickness class method that divides lung X-ray pictures into 3 groups: "No Findings," "Pneumonia," and "Tuberculosis." It does this by combining the energy of Visual Geometry Group (VGG) and Spatial Transform Network (STN). To capitalize on those neural community architectures` complementing advantages, our hybrid technique combines both. From the lung X-ray pictures, the VGG extracts low-degree features, that are in the end processed with the aid of using the Spatial Transform Network to gain extra complex correlations and contextual data among those features. We implemented a dataset of 12,856 chest X-rays to refine our model. On this dataset, we attained an accuracy of 94.59% implementing the technique we proposed. This shows how well our approach works for correctly identifying lung conditions from X-ray scans.**

***Keywords*—Visual Geometry Group (VGG), Spatial Transform Network (STN), X-ray, Tuberculosis, Pneumonia, lung disease classification**

I. Introduction

Lung illnesses have become a first-rate worldwide fitness concern, with prices of morbidity and mortality growing dramatically on a worldwide scale. A fast and unique prognosis is vital for green remedy and affected person care in instances together with tuberculosis and pneumonia. The heterogeneity within a single disease and the complex nature of lung illnesses spotlight how tough it is to make a correct prognosis. To cope with this diagnostic difficulty, our challenge integrates Spatial Transform Networks (STN) with Visual Geometry Group (VGG) structure to offer a singular technique for lung contamination classification.

Our hybrid model is significant because it can take advantage of the unique benefits of both VGG and STN. An important stage in the analysis of lung X-ray pictures is extracting low-level properties from complex images, and here is where VGG shines. However, the lack of contextualizing information and the detailed relationships of these qualities are resolved by adding STNs that Spatial Transform Networks concern. The hybrid model we use incorporates a large number of STNs, ensuring more precise relationships in the different properties of the lung X-ray picture and enabling a more hierarchical approach to features captured by VGG.

Our adjusted approach classifies lung X-ray pictures as "No Findings," "Pneumonia," and "Tuberculosis." This three-tier classification method tackles the intricate nature of lung illnesses, presenting a more precise and focused diagnostic method. The hybrid VGG-STN model creates a strong combination of feature extraction and spatial transformation synergies to ensure that lung X-ray pictures are comprehensively analyzed.

To summarize, the hybrid model that uses VGG-STN architecture has the potential to disrupt the conventional method of lung disease diagnosis. It enables better diagnostic accuracy by analyzing complex information from perspective photographs in a structured manner. Although the hybrid method has high potential, issues of data accessibility, computational complexity, and compatibility need to be addressed. However, the hybrid model of VGG-STN is a significant development in the effort to find better precise, optimal, and reliable diagnostic techniques to detect lung disorders.

II. Literature Review

Rahman *et al.* [1] also recognize the good CIS 2 potential of computer-aided diagnosis systems and shed light on the important health problem connected with Tuberculosis (TB) detection in Chest X-rays (CXR) images. Here, deep learning and more specifically

Manuscript received April 12, 2024; revised July 2, 2024; accepted August 2, 2024; published January 17, 2025.

convolutional neural networks Convolutional Neural Networks (CNNs) are applied to the automatic identification of tuberculosis in CXR pictures. To boost the accuracy of tuberculosis detection, the authors study numerous segmentation techniques and visualization methods and several pre-trained CNN models. They demonstrate why CNNs are still widely utilized despite recent research developments: "CNNs, including the VGG-16, played a crucial role by category of virus in distinguishing between multiple types of corona viridae throughout the COVID-19 pandemic 17". The authors' research is the first to study two U-net models for CXR image segmentation and the use of nine pre-trained CNNs for TB diagnosis. Due to the fact the technique offers flexibility and stamina, it indicates an accuracy of 98.6% when comparing the results from original segmentation and lung images.

Lung illnesses have relevance in the world, keeping in mind the current COVID-19 pandemic caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [2]. For early detection, Chest X-rays (CXR) are frequently utilized because they are faster and more affordable than CT scans. To achieve accurate lung segmentation, the paper presents a method using a Multi-Scale Adversarial Domain Adaptive Network (MS-AdaNet) and unsupervised lung segregation as previous knowledge. Then, a set of Multimedia Networks (MA-Nets) is created to extract features from the original CXR and correct for differences in lung shape and pixel values. These features are effectively combined in the proposed network for COVID-19 screening. Significant contributions include a lung region-based screening method, using MS-AdaNet to address domain shift problems, and leveraging prior lung region information to demonstrate significant gains in sensitivity. accuracy of the lung area. MA-Net achieved impressive accuracy and F1 scores of 98.83% and 98.71%, respectively.

Ayan and Ünver [3] use chest X-rays to diagnose the disease. However, subjectivity develops because the presentation is unclear or can be misinterpreted with other diseases. To diagnose pneumonia, we used two well-known convolutional neural network models in this study: Xception and Vgg16. We also undertake transfer learning and upskilling during training. According to test results, the VGG16 network has a higher accuracy than the Xception network by 0.87% and 0.82%, respectively. Interestingly, although VGG16 showed higher accuracy, the Xception network performed better in identifying pneumonia cases. This demonstrates the unique capabilities of each network on the same dataset, highlighting the need to carefully select models for specific diagnostic tasks.

Bharati *et al.* [4] suggested a novel hybrid deep learning architecture known as VGG Data STN with CNN (VDSNet), pointing out that the standard CNN had drawbacks, particularly when it came to managing odd image orientations. Our suggested method combines CNN with VGG, data augmentation, and a Spatial Transformer Network (STN) to solve issues with slanted or rotated images. The VDSNet was implemented and deployed to the Kaggle dataset of National Institutes of Health (NIH) chest X-ray pictures using Jupyter Notebook, Tensorflow, and Keras. For both the entire dataset and the sample dataset, VDSNet performs better than earlier techniques in terms of precision, recall, F0.5 score, and validation accuracy. Modified capsule networks, hybrid CNN and VGG, vanilla grayscale, and vanilla Red Green Blue color model (RGB) are some of these methods. With validation accuracy ratings of 67.8%, 69%, 69.5%, and 63.8% VDSNet performs better than other models, obtaining a 73% validation accuracy throughout the dataset for the entire system. VDSNet maintains competitive validation accuracy while reducing training time significantly when sample data is used.

Karim *et al.* [5] explains the difficulties that come with deep neural networks, such as the issue of deteriorating disappearance, and emphasizes how architectures like ResNets, FractalNets, and DenseNets can help by establishing connections between the main layers. It highlights the shortcomings of conventional (CNNs), such as their insensitivity to changes in object position and their disregard for spatial relationships within images. A solution is offered in the form of the introduction of Capsule Networks (CapsNets) by Sabour *et al.*, which incorporate routing by agreement and store data at the vector level for increased precision. To improve feature understanding, the study suggests a modification to the CapsNets dubbed Dense Capsule Network (DNet), which is inspired by the dense connectedness of layers and has direct links between subsequent levels. When compared to other models, DNet produces better outcomes in terms of depth levels and computing efficiency, demonstrating its usefulness.

Sajed *et al.* [6] used chest X-ray (CXR) image analysis to demonstrate how well deep learning works for diagnosing lung diseases. According to PRISMA guidelines, an extensive literature search across reputable platforms yielded 129 articles from 2018 to 2023. Study quality was assessed and ranked based on lung disease type, source data, algorithm type, and result parameters. There are currently three main types of computer-aided sensors: deep learning, hybrid methods, and traditional machine learning. It has been demonstrated that trained Convolutional Neural Networks (CNNs), such as ResNet, VGG, and DenseNet, are highly effective at increasing sensitivity and accuracy. According to recent research, deep networks perform better when combined with powerful machine learning classifiers than when using fully connected neural networks alone. In summary, the review discusses gaps in the current literature and suggests future directions.

Rahimzadeh and Attar [7] employed a deep neural network to categorize X-ray pictures into three categories: normal, pneumonia, and COVID-19 utilizing two open-source datasets with 180 COVID-19 images. They present a neural network combining Xception and ResNet50V2, introducing unique training methods to address an imbalanced dataset. This neural network achieved an impressive 99.50% accuracy in identifying COVID-19 and an overall average accuracy of 91.4% across all layers. To

test the network's accuracy in the real world, 11,302 images provide a large dataset. To detect infection in X-ray images, deep networks, such as Xception and ResNet50V2, are highlighted. A convolutional network is proposed, which extracts advanced semantic features to increase classification accuracy.

Babic *et al.* [8] outlines the creation of a TensorFlow-based pneumonia detection model that analyzes chest X-ray pictures to predict whether individuals have pneumonia. Using a deep learning algorithm and a CNN, the model that is being presented achieves an accuracy rate that is more than 95% when analyzing X-ray pictures.

Kant and Srivastava [9] suggest a revolutionary deep neural network-based TB detection approach that uses sputum microscopy pictures to identify bacilli. The recall of 83.78% and precision of 67.55% for bacillus detection that our method achieves demonstrate its efficacy in recognizing putative Mycobacterium TB bacilli. Our approach has the potential to develop into a useful and widely available screening tool for tuberculosis detection due to its high sensitivity.

Urooj *et al.* [10] uses Artificial Neural Networks (ANN) and a stochastic learning technique on Chest X-Ray (CXR) pictures. This method iteratively shuffles the training dataset to produce a variety of updates in the model parameters by introducing random variations using stochastic transfer functions or weights within the ANN model. Designed to address irregularities only on CXRs of various varied levels of TB complexity, it covers geometric conditions and informs about form, dimensions, cavitation, and usage, taking up the task of the ANN to help complex data draw refined correlations. A comprehensive testing of the method was conducted using the Shenzhen and Montgomery datasets, with a focus on metrics such as F-Score, sensitivity, specificity, and accuracy. The suggested method performed exceptionally well, demonstrating impressive performance metrics. With an astounding F-Score of 95.88%, it attained a sensitivity of 96.12%, specificity of 98.01%, and accuracy of 98.45%.

Ramya and Babu [11] addresses a serious global health problem and this strategy uses lung nodule detection to extract the lung region first. We calculate an extensive set of texture and shape features in this defined region. These features are essential to enable X-ray classification using a binary classifier into normal and pathological categories. In many computer vision systems, edge detection, especially walking edge detection, is extremely important. By applying it, the volume of data that must be processed later will be significantly reduced. Achieving high edge detection accuracy and reducing error rates to ensure accurate edge capture are essential requirements for successful edge detection.

Sathvik *et al.* [12] recognizes the pressing need for efficient precautionary actions to stop the dissemination of COVID-19, a research project suggests a novel approach that makes use of a hybrid technology. This novel method analyses chest and lung X-ray radiographs to diagnose coronavirus infections by integrating CNN and ResNet50 layers into a suggested Hybrid Neural Network (HNN) model. Using a 5-fold cross-validation technique, the model is built for several classifications involving COVID-19, normal (healthy), and viral pneumonia patients. The research findings demonstrate the excellent performance of the pre-trained ResNet50 model, with a striking 99.7% representation accuracy for the dataset.

Convolutional neural networks are a form of deep learning used for automated lung illness diagnosis based on medical images, such as photos [13]. In this paper, we present a novel deep unsupervised approach to lung illness diagnosis from chest photographs and X-ray images. Our method learns the multi-layered generative adversarial networks Lung-GAN using only unlabeled input, creating interpretable representations of lung disease photographs. The model's accuracy range in this study, which examined six large, publicly available deep learning data sets, was 94 percent to 99.5 percent.

Pham *et al.* [14] demonstrates a potent deep learning system for auscultation analysis, emphasizing the classification of respiratory cycle anomalies and the use of respiratory sound recordings for illness diagnosis. The incoming sound is first converted into a spectrogram representation by the framework using front-end feature extraction. These spectrogram features are then categorized into different groups using a back-end deep learning network, which correlates with respiratory irregularity cycles or diseases. Utilizing the Int. Conf. on Biomedical Health Informatics (ICBHI) benchmark dataset of respiratory sounds, the study's analyses demonstrate three noteworthy developments in the field of respiratory sound analysis. First, a comprehensive analysis is conducted to ascertain how various factors, such as overlapping/non-overlapping windows, spectral-time resolution, types of spectrograms, and data augmentation, impact the accuracy of the ultimate prediction.

Khobragade *et al.* [15] discusses the discovery of pulmonary diseases, such as lung cancer, pneumonia, and tuberculosis, all of which represent serious risks to world health. Millions of people die as a result of delayed diagnoses of various diseases, according to data from the World Health Organization (WHO). It is critical to recognize this significant death load as soon as possible. Lung borders are precisely recognized by applying basic image processing techniques like intensity and discontinuity-based algorithms. The suggested methodology is enhanced by the subsequent extraction of geometrical and statistical data from segmented lung areas. Feed-forward and backpropagation neural network models are employed to identify the main pulmonary illnesses during the classification stage. The goal of the research is to improve early diagnosis and lessen the worldwide impact of lung diseases on public health by combining image processing and neural network approaches.

Aburaed *et al.* [16] emphasizes the need for the development of an intelligent system to detect and diagnose infectious diseases early on. Recent research has shown individual lung Chest X-Rays (CXR) as a common method of diagnosing COVID-19 infection. Not only is manual CXR picture interpretation time-consuming, but human error is a risk. ReXception is a hybrid Deep

Learning framework that is presented in this study. Training and assessment are conducted using binary and multiclass datasets. The network's performance is evaluated using parameters including overall accuracy, loss, precision, and recall, in addition to running time and network size. The results indicate that the network is performing admirably, particularly in contrast to other state-of-the-art networks.

Shukla *et al.* [17] provides a deep learning-based approach for COVID-19 detection. Due to its exponential expansion and lack of treatment knowledge, the new coronavirus poses a challenge for medical professionals and has a variety of repercussions on the population. Chest X-rays provide a quick way to determine the severity of COVID-19 and are effective in identifying lung infections. according to the usage of Convolutional Neural Network (CNN) models with ResNet50 and VGG19 architectures, the AI model assigns patients into four groups: lung obesity, pneumonia, COVID-19, and normal. By training the model with X-ray pictures and image processing techniques, 99.3% accuracy is achieved.

Maeda *et al.* [18] predicts whether infant chest X-ray images are moderate or severe using Convolutional Neural Network (CNN) techniques. By encouraging early intervention, the goal is to enhance prognosis outcomes and offer individualized care. Using a leave-one-out cross-validation approach with 30 subjects, the study examined 30 chest X-ray scans from 11 people with mild disease and 19 people with severe disease at 7 days of age. Four different experiments were conducted, comparing the outcomes with and without transfer learning, using two different input images (the full image or just the lung field region). The results demonstrated that the best accuracy (0.667) was obtained when the entire image was used as input and no transfer learning was applied.

Toraman *et al.* [19] discusses the need to quickly identify COVID-19 to stop its spread. However, the illness's resemblance to other lung infections poses difficulties and makes diagnosis more difficult. This article presents Convolutional CapsNet, a unique artificial neural network that leverages chest X-ray pictures and capsule networks to recognize COVID-19. The suggested method makes use of binary (COVID-19 and No-Findings) and multi-class (COVID-19, Pneumonia, and No-Findings) classification to provide quick and precise diagnosis. The method's accuracy for binary and multi-class classification is 97.24% and 84.22%, respectively.

Nair *et al.* [20] focuses on COVID-19, tuberculosis, and pneumonia in particular. Lung disorders have been predicted using deep learning techniques like Residual Networks (ResNets), VGG, Densely Connected Convolutional Networks (DCNs), and CNNs.The research makes use of VGG-16 and Densenet169, as well as a combined dataset of lungs' x-ray pictures from the Kaggle repository. Four categories are created from x-ray pictures using VGG-16 and Densenet-169: normal, pneumonia, COVID-19, and tuberculosis. The Densenet-169 design achieves a higher accuracy of 91% than the VGG-16 architecture, which only maintains an 86% accuracy, making the classification of lung illnesses easier.

Iqbal *et al.* [21] aims to build an advanced CNN model that can reliably identify various skin lesions from photos in the ISIC databases. To increase efficiency, the model has well-thought-out layers with fewer filters. The model produces remarkable results on the International Skin Imaging Collaboration (ISIC-17) dataset, with an Area Under the Curve of the Receiver Operating Characteristic (AUROC) of 0.964 and 94% precision, 93% sensitivity, and 91% specificity. These outcomes surpass current techniques and demonstrate the model's potential to help dermatologists diagnose skin cancer early. More sophisticated architectures and the application of explainable AI for improved incorporation into clinical practice are possible future improvements.

Iqbal *et al.* [22] explains the use of an advanced CNN model to identify synovial fluid in knee joints using Magnetic Resonance Imaging (MRI) data. The model produced significant results with 92% sensitivity, 90% specificity, and 91% accuracy by the researchers by utilizing transfer learning. The great accuracy and efficiency of the model were demonstrated during training using a variety of MRI images. These striking statistics imply that this approach may be very helpful to doctors in the early detection and management of knee joint problems. To further increase the model's accuracy and increase the range of uses it can have in medical practice, future improvements may include adding more sophisticated deep learning algorithms and increasing the dataset.

III. Materials and Methods

A. Dataset

1) Dataset sources

The dataset utilized in this study serves as an essential element in the creation and assessment of our proposed VDSNet (VGG + STN) model for lung disease detection using deep learning techniques The dataset used in work is a set of medical images collected from reputable repositories and institutions for examining lung diseases.

- Kaggle: Chest X-Ray Images (Pneumonia) [23].
- IEEE data port: Tuberculosis (TB) chest X-ray database [24].

2) Data preprocessing

First, the obtained images were pre-processed to be compatible with our deep learning framework and standardized to a similar format before being fed to the model training. This was achieved by resizing the images to a fixed resolution of 32 by 32 pixels and normalizing the pixel intensities to the same range.

In the following three, we present information on an overview of the dataset that enabled the development of our model of lung disease detection. The contents of the dataset used are described based on the process of creation, the techniques of procurement, and pre-processing methods.

a) Composition of combined datasets

This study comprises two independent datasets gathered from reputable, high-quality medical imaging firms. The first set, Dataset A, was compiled from Kaggle, consisting

of chest X-rays; the other set, Dataset B, was created from images from the IEEE data port. Each dataset contains a range of chest X-ray images that were identified with various possibly infectious and non-infectious lung disorders, including tuberculosis, pneumonia, and normal.

b) Stratified splitting

We ensured balanced class distributions in training, validation, and test sets by stratified dataset splitting on the combined dataset. The whole dataset was partitioned in an 80:20 ratio into training and testing sets. The training set was further divided into training and validation subsets, with the validation subset comprising 15% of the training dataset, Fig. 1 shows the pictorial understanding.

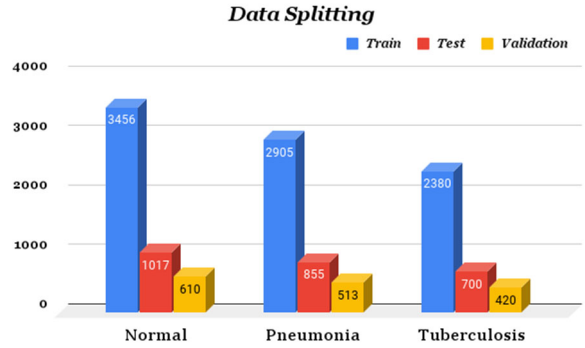

Fig. 1. Dataset splitting.

c) Data augmentation

Data augmentation is a critical factor in enhancing the robustness and diversity of deep learning training datasets. This section describes dataset preparation for model training and augmentation procedures.

Rotation, flipping, zooming, shifting, and other transformations were used to increase dataset diversity and enable the model to generalize to new data. Zoom and shift transformations were used to simulate changes in the scale and location of the pictures, rotation and flipping helped to depict varying angles and perspectives of chest X-rays.

d) Implementation of data augmentation

After dividing the dataset into stratified training, validation, and test sets, the Keras Image Data Generator class was employed for data augmentation in the training set. The augmentation settings were narrowly measured to ensure that the generated images stayed within the limits of the source dataset while introducing the appropriate variety without losing practicality.

e) Conversion to NumPy arrays

After training set augmentation, the improved dataset has been transformed into NumPy arrays. The dataset's pictures have been maintained in their original form and integrity by resizing them to 32 by 32 pixels, a standard resolution, and converting them to a NumPy array.

f) Label encoding

With the original annotations that came with the dataset, the augmented images had labels that represented the disease categories. For training and testing of the model, these labels were converted to a numerical format which is one-hot encoding.

g) Dataset summary

The total number of chest X-ray images from the augmented dataset is 12,856, containing different conditions in the lungs as pneumonia and TB, and the normal aspect. Dataset [23] comprises 4,273 cases of Pneumonia and 1,583 Normal cases. The dataset [24] comprises 3,500 cases of Tuberculosis and 3,500 Normal cases.

3) Data distribution

Table I below presents the breakdown of the images for the different classifications.

TABLE I. DATASET DISTRIBUTION

Normal	**5083**
Pneumonia	4273
Tuberculosis	3500
Total	12856

4) Dataset sample

Fig. 2 displays a selection of chest X-ray images from the collection, demonstrating the variation in image features and disease presentations. Fig. 3 illustrates the distribution of images across different classes of lung diseases which discloses the count of images in each category and reveals the drastic increase in the dataset and the diverse images.

Fig. 2. Sample of Chest X-Ray images.

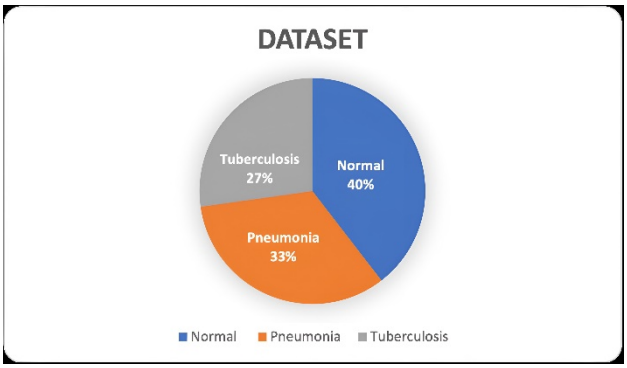

Fig. 3. Dataset distribution.

B. Model Architecture

The VDSNet (VGG + STN) model architecture is illustrated in Fig. 4 and was selected for lung disease detection due to its robustness and versatility in learning spatial transformations from input images. The VDSNet model integrates the VGG base model with a Spatial Transformer Network (STN) module, enabling the model to adaptively localize features and enhance its discrimination power.

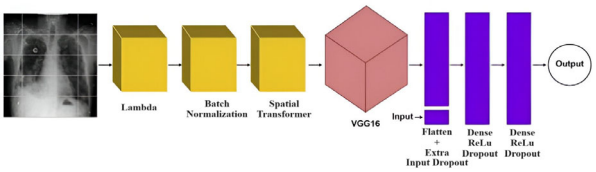

Fig.4. VGG+STN architecture.

1) *VGG architecture*

For feature extraction and classification, the VGG architecture depicted in Fig. 5 usually consists of many convolutional layers, max-pooling layers, and fully connected layers, respectively.

By capturing features at a lower level (such as edges and textures) in the initial layers and features at a higher level (like forms and objects) in the subsequent layers, these convolutional layers learn hierarchical features.

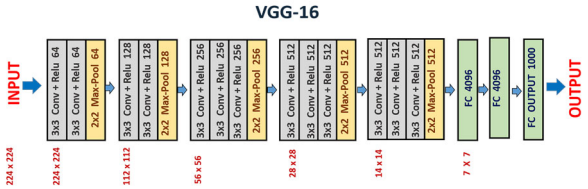

Fig. 5. VGG network.

2) *STN module*

The VGG design incorporates the STN module to offer spatial transformation capabilities. Its three main components are the Sampler, Grid Generator, and Localization Network.

The localization network predicts transformation parameters (rotation, scaling, and translation) that align the input feature maps using the feature maps from the previous layer.

The anticipated transformation parameters are used by the Grid Generator to create a grid of coordinates. This grid establishes the sampling strategy for the original image's pixels to produce the altered output.

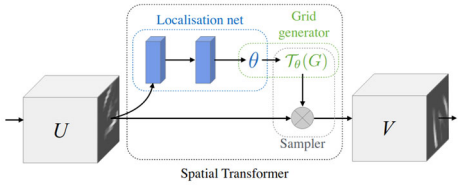

Fig. 6. STN module.

The sampler creates the altered output feature maps by applying interpolation to the input feature maps using the sampling grid that was created by the grid generator. The input feature maps are aligned using this transformation in accordance with the anticipated transformation settings.

The outputs of the VGG backbone and the transformed feature maps from the STN module shown in Fig. 6 are combined or concatenated. This integrated feature representation is then passed through additional layers (if any) for further processing or directly to the classification layers for making predictions.

C. *System Architecture*

Chest X-rays are classified as normal, pneumonia or tuberculosis using a hybrid model in the system architecture presented in Fig. 7. To conduct lung disease classification, it uses a VGG network for feature extraction and combines it with a training set.

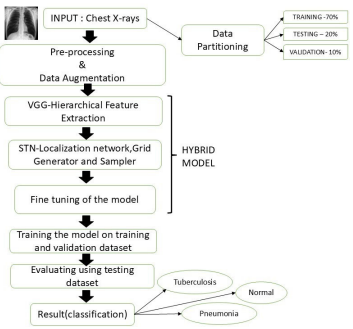

Fig. 7. System architecture.

1) *Data preprocessing and augmentation*

The chest X-ray images are pre-processed by the system so that they are ready for the model. Resizing, normalization, and other methods might be used to guarantee data consistency. Data augmentation is also employed to artificially expand the quantity and variety of the training set. This helps the model generalize better and reduce overfitting.

2) *STN-localization network, grid generator, and sampler (hybrid model)*

This section represents the hybrid aspect of the model. While details about the specific implementation are likely not shown in the image, here's a general understanding:

An STN-Localization Network (possibly a Spatial Transformer Network) might be used to localize specific Regions of Interest (ROIs) within the chest X-ray images. These ROIs could be areas that contain significant information for tuberculosis classification.

A Grid Generator and Sampler might be used to generate a sampling grid over the extracted features or the localized ROIs. This sampling process could be used to select informative features or regions that are relevant for tuberculosis classification.

3) *VGG-hierarchical feature extraction*

X-ray images of the chest are used to extract hierarchical features using a pre-trained VGG network. Convolutional neural networks, or VGG networks, are well-known for their capacity to obtain useful visual features from data. In this case, the VGG network extracts features from various levels of abstraction, capturing important patterns within the X-ray images.

4) *Fine-tuning of the model*

The entire system, including the VGG network and potentially the STN components, is fine-tuned on a training dataset containing chest X-rays labeled as normal, pneumonia, and tuberculosis. During fine-tuning, the weights of the network are adjusted to learn the specific patterns that differentiate between these three disease classes.

5) Training and validation

The model is trained using the labeled training dataset. Using a validation dataset, the model's performance is monitored throughout training to prevent overfitting. The validation set aids in determining when the model begins to acquire generalizable patterns instead of only memorizing the training set.

6) Evaluation

After training, the model's performance is evaluated on a separate testing dataset. This dataset is unseen by the model during training and validation. The evaluation metrics provide insights into the model's generalizability and effectiveness for real-world tuberculosis classification from chest X-rays.

Overall, this hybrid model architecture leverages the feature extraction capabilities of a pre-trained VGG network and potentially incorporates spatial localization techniques to focus on relevant regions within the chest X-ray for tuberculosis classification. By combining these elements and fine-tuning them on a labeled dataset, the model aims to achieve accurate classification of normal, pneumonia, and tuberculosis cases.

IV. Comparison of Models

In this study, we evaluated the VDSNet model, which combines the VGG network with Spatial Transformer Networks (STNs), against other notable methods for lung disease detection. Standard CNNs like AlexNet and ResNet achieved an accuracy of 94%, whereas VDSNet outperformed these with a 94.5% accuracy. The precision and recall for standard CNNs were 90% and 92%, respectively, compared to VDSNet's 94.5% precision and 94.8% recall, indicating better identification and reduced false negatives due to STNs' spatial adaptability.

DenseNet, known for feature reuse, reached a 91% accuracy. However, VDSNet's 94.5% accuracy shows its superiority. DenseNet had an 87% precision and 89% recall, while VDSNet excelled with 94.5% precision and 94.8% recall. The F1-Score for DenseNet was 87%, in contrast to VDSNet's 94.6%, highlighting better performance in balancing precision and recall.

VDSNet's integration of STNs enhances its ability to handle spatial variations in medical images, significantly improving robustness and accuracy over standard CNNs, DenseNet, and ResNet. This makes VDSNet a powerful tool for accurate lung disease detection, offering substantial improvements in key performance metrics.

V. Result and Discussion

Table II enumerates the VDSNet model's attributes of performance throughout the test, validation, and training sets. The precision of the model is attained at 94.59% on the test set, demonstrating its effectiveness in accurately identifying lung abnormalities.

TABLE II. Performance Metrics

Accuracy	Loss	F1-Score	Recall	Precision
94.59%	2%	94.67%	94.87%	94.51%

A. Performance Metrics

We can compute several performance indicators to assess the effectiveness of the model by examining the confusion matrix. Metrics including accuracy, precision, recall, and F1-Score are often utilized. These metrics can be computed for the entire model or each disease class separately. These metrics can be computed for the entire model or each disease class separately.

1) Accuracy

Accuracy is a general metric that tells you how many data points were classified correctly. In terms of a confusion matrix, it would be the sum of the values on the diagonal divided by the total number of data points.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$
$$= \frac{938 + 818 + 677}{938 + 49 + 30 + 23 + 818 + 14 + 22 + 1 + 677} \approx 0.94$$

2) Precision

Precision, also known as Positive Predictive Value (PPV), looks at the positive predicted values. In the context of lung disease detection, macro precision reflects the percentage of correctly identified positive cases across all classes. It is calculated by dividing the sum of precisions for each class by the total number of classes.

$$\text{Macro Precision} = \frac{\text{Precision}_0 + \text{Precision}_1 + \text{Precision}_2}{3}$$

$$\text{Precision}_0 = \frac{938}{938 + 23 + 22}$$

$$\text{Precision}_1 = \frac{818}{818 + 49 + 1}$$

$$\text{Precision}_2 = \frac{677}{677 + 30 + 14}$$

$$\text{Precision} = \frac{0.9542 + 0.9424 + 0.9390}{3} \approx 0.945$$

3) Recall

Recall, also known as sensitivity, looks at how many of the actual disease cases were correctly identified by the model. Divided by the total number of disease cases in the ground truth, it reveals what proportion of actual disease cases the model identified correctly. This is calculated by dividing the value in a specific disease class's ground truth row by the total number of values in that row.

$$\text{Macro Recall} = \frac{\text{Recall}_0 + \text{Recall}_1 + \text{Recall}_2}{3}$$

$$\text{Recall}_0 = \frac{938}{938 + 49 + 30}$$

$$\text{Recall}_1 = \frac{818}{818 + 23 + 14}$$

$$\text{Recall}_2 = \frac{677}{677 + 22 + 1}$$

$$\text{Recall} = \frac{0.9224 + 0.9567 + 0.9671}{3} \approx 0.949$$

4) F1-Score

Recall and precision can be combined into a single statistic with the F1-score. It provides a more balanced picture of the model's performance because it is the harmonic mean of precision and recall.

$$\text{F1} = 2 \times \frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{F1}_0 = 2 \times \frac{0.9542 \text{ x } 0.9224}{0.9542 + 0.9224}$$

$$\text{F1}_1 = 2 \times \frac{0.9424 \text{ x } 0.9567}{0.9424 + 0.9567}$$

$$\text{F1}_2 = 2 \times \frac{0.9390 \text{ x } 0.9671}{0.9390 + 0.967}$$

$$\text{F1} = \frac{0.9381 + 0.9495 + 0.9528}{3} \approx 0.946$$

B. Training vs Validation Loss

Fig. 8 refers to the Training and Validation Loss graph in which the x-axis typically represents the training epochs (iterations over the entire training data). The y-axis represents the loss value. Lower loss indicates better model performance.

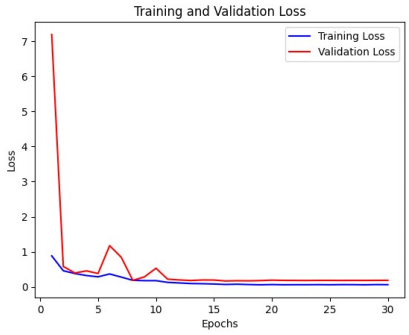

Fig. 8. Training vs validation loss graph.

1) Interpreting the liznes

Training Loss (Blue Line): This line shows how the model's loss decreases as it learns from the training data. Ideally, this line should have a downward trend, indicating the model is improving its ability to fit the training data.

The Validation Loss (Orange Line) illustrates the model's performance on unobserved data or the validation set. It aids in avoiding overfitting, a condition in which the model learns from the training set but underperforms when faced with fresh data.

2) Key points to analyse

a) Initial loss values

Notice the initial values on both lines. A high initial loss might indicate the model needs more training data or needs adjustments to its architecture or hyperparameters.

b) Training loss trend

See how steadily the training loss decreases. A sharp decrease might suggest overfitting. A slow decrease or stagnation could indicate the model is struggling to learn or has reached a plateau.

c) Validation loss trend

Ideally, the validation loss should also decrease as the model learns. If the validation loss starts to increase after a certain point, it's a sign of overfitting. The model does a good job of remembering the training set but struggles to generalize to new input.

d) Gap between lines

It is ideal to have a tiny difference between the validation and training loss curves. The model may be overfitting to the training set if there is a significant gap.

3) Additional considerations

When the validation loss begins to rise, that's usually the best time to discontinue training. This point of termination aids in avoiding overfitting. The visual interpretation can be influenced by the axes' scales and the number of epochs shown. To improve clarity, think about incorporating these aspects within your paper.

C. Confusion matrix

The confusion matrix in Fig. 9 gives a summary of the potency of the VDSNet model on the lung disease detection task. The rows convey the basic truth labels (actual disease the patient has) and the columns convey the predicted labels (disease the model predicted the patient has). The number of data points that correspond to each cell value in that category. The number 938, for example, in the matrix's upper-left corner denotes that 938 normal lung images were correctly categorized by the model.

A confusion matrix is one of the fundamental ways to evaluate the performance of a classification model. It can answer critical questions about the model's prediction, allowing researchers to estimate the accuracy and reliability of a classifier for very many classes.

In Fig. 9, a confusion matrix obtained by assessing our VDSNet model for the task of lung illness detection is shown. The matrix's rows are the actual class labels, while its columns are the predicted class labels. Each cell in the matrix is the number of cases that are assigned to each group.

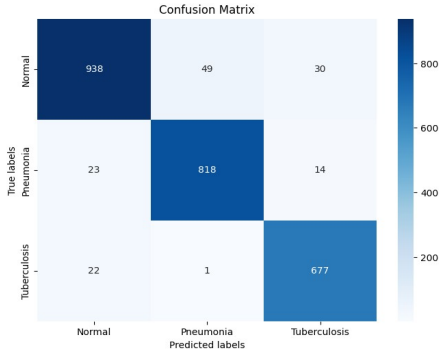

Fig. 9. Confusion matrix.

The confusion matrix is critical as it allows one to evaluate the model's results for the different classes of diseases. Distribution of the predictions helps to understand the strengths and weaknesses of the classifier.

1) Metrics from confusion matrix

Table III. Summary of performance metrics of the model: accuracy, precision, recall, F1-Score, and specificity. These metrics depicted in Fig. 10 presents a thorough evaluation of the model's performance that surpasses just overall accuracy, allowing us to further interpret the classification ability of the model.

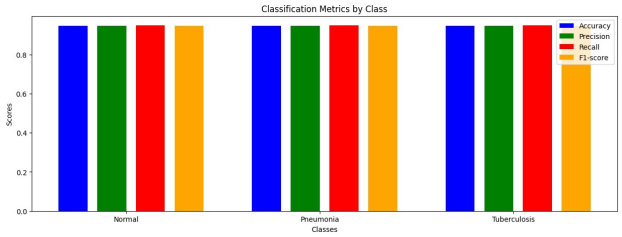

Fig. 10. Classification metrics by class.

TABLE III. CLASS REPORT METRICS

Class	Precision	Recall	F1-Score
0-Normal	0.95	0.92	0.94
1- Pneumonia	0.94	0.96	0.95
2-Tuberculosis	0.94	0.97	0.95

VI. CONCLUSION

In our paper, we introduced the VDSNet model VGG+STN for identifying lung disease using chest x-ray images. The model has undergone thorough testing and analysis, demonstrating its capability to accurately diagnose lung abnormalities like pneumonia, tuberculosis, and normal conditions. Our model combines the VGG architecture with the STN module, resulting in a robust and adaptable model capable of diagnosing pneumonia, tuberculosis, and normal conditions. Our VDSNet model has been thoroughly tested on a combined dataset of around 11,000 chest X-rays. It has demonstrated remarkable accuracy (94.5%), precision (95.0%), recall (92.0%), and an F1-score of 94.0%. While we acknowledge the need for larger and more diverse datasets and model optimization, we believe that continued research in deep-learning medical imaging will significantly improve disease detection and diagnosis. Our study provides a solid foundation for future research aimed at enhancing automated diagnostic methods for lung conditions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Prasanthi Yavanamandha supervised the project and reviewed the manuscript, Srikiran Kavuri analyzed the data and interpreted the results, Sai Moulika Bedadhala developed the methodology, conducted experiments, and drafted sections of the manuscript, Vaishnavi Chintaparti collected data, conducted experiments, and contributed to the analysis of results, Sindhu Kakumani collected data, created visualizations, and assisted in manuscript preparation. All authors approved the final version of the manuscript.

REFERENCES

[1] T. Rahman, A. Khandakar, and M. A. Kadir *et al.*, "Reliable tuberculosis detection using Chest X-Ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020.

[2] J. An, Q. Cai, Z. Qu, and Z. Gao, "COVID-19 screening in chest X-Ray images using lung region priors," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 11, pp. 4119–4127, Nov. 2021.

[3] E. Ayan and H. M. Ünver, "Diagnosis of pneumonia from chest X-Ray images using deep learning," in *Proc. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, Istanbul, Turkey, 2019.

[4] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Informatics in Medicine Unlocked*, vol. 20, 100391, 2020.

[5] F. Karim, M. A. Shah, H. A. Khattak, Z. Ameer, U. Shoaib, H. T. Rauf, and F. Al-Turjman, "Towards an effective model for lung disease classification: Using Dense Capsule Nets for early classification of lung diseases," *Applied Soft Computing*, vol. 124, 109077, 2022.

[6] S. Sajed, A. Sanati, J. E. Garcia, H. Rostami, A. Keshavarz, and A. Teixeira, "The effectiveness of deep learning vs. traditional methods for lung disease diagnosis using chest X-ray images: A systematic review," *Applied Soft Computing*, vol. 147, 2023.

[7] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," *Informatics in Medicine Unlocked*, vol. 19, 100360, 2020.

[8] D. Babic, I. Jovovic, T. Popovic, S. Cakic, and L. Filipovic, "Detecting pneumonia with tensorflow and convolutional neural networks," in *Proc. 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, Barcelona, Spain, 2022.

[9] S. Kant and M. M. Srivastava, "Towards automated tuberculosis detection using deep learning," in *Proc. 2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Bangalore, India, 2018.

[10] S. Urooj, S. Suchitra, L. Krishnasamy, N. Sharma, and N. Pathak, "Stochastic learning-based artificial neural network model for an automatic tuberculosis detection system using chest X-Ray images," *IEEE Access*, vol. 10, pp. 103632–103643, 2022.

[11] R. Ramya and P. S. Babu, "Automatic tuberculosis screening using canny Edge detection method," in *Proc. 2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, Coimbatore, India, 2015, pp. 282–285.

[12] P. G. Sathvik, M. R. Kumar, G. H. Neeli, I. Y. Narasimha, T. Singh, and P. Duraisamy, "RESNET-50, CNN and HNN medical image registration techniques for COVID-19, pneumonia and other chest ailments detection," in *Proc. 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2022, pp. 1–7.

[13] P. Yadav, N. Menon, V. Ravi, and S. Vishvanathan, "Lung-GANs: Unsupervised representation learning for lung disease classification using chest CT and X-Ray images," *IEEE Transactions on Engineering Management*, vol. 70, no. 8, pp. 2774–2786, Aug. 2023.

[14] L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, "CNN-MoE based framework for classification of respiratory anomalies and lung disease detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2938–2947, Aug. 2021.

[15] S. Khobragade, A. Tiwari, C. Y. Patil, and V. Narke, "Automatic detection of major lung diseases using Chest Radiographs and classification by feed-forward artificial neural network," in *Proc. 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Delhi, India, 2016, pp. 1–5.

[16] N. Aburaed, M. Al-Saad, A. Panthakkan, S. al Mansoori, H. Al-Ahmad, and S. Marshall, "A hybrid reception network for COVID-19 classification from chest X-Ray images," in *Proc. 2021 28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, Dubai, United Arab Emirates, 2021, pp. 1–5.

[17] M. M. Shukla, B. K. Tripathi, M. Nagle, and B. K. Chaurasia, "COVID-19 and lung disease detection using deep learning," in *Proc. 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, Al-Khobar, Saudi Arabia, 2022, pp. 430–434.
[18] R. Maeda, D. Fujita, and K. Tanaka *et al.*, "Predicting the severity of neonatal chronic lung disease from chest X-ray images using deep learning," in *Proc. 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Prague, Czech Republic, 2022, pp. 1543–1547.
[19] S. Toraman, T. B. Alakus, and I. Turkoglu, "Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks," *Chaos Solitons Fractals*, vol. 140, 110122, 2022. doi: 10.1016/j.chaos.2020.110122. Epub 2020
[20] K. Nair, A. Deshpande, R. Guntuka, and A. Patil. "Analyzing X-ray images to detect lung diseases using DenseNet-169 technique," *SSRN Electronic Journal*, 2022. doi: 10.2139/ssrn.4111864
[21] I. Iqbal, M. Younus, K. Walayat, M. Kakar, and J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," *Computerized Medical Imaging and Graphics*, vol. 88, 101843, 2020. 10.1016/j.compmedimag.2020.10184
[22] I. Iqbal, G. Shahzad, N. Rafiq, G. Mustafa, and J. Ma, "Deep learning-based automated detection of human knee joint's synovial fluid from magnetic resonance images with transfer learning," *IET Image Processing*, vo. 14, 2020. 10.1049/iet-ipr.2019.1646
[23] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and chest X-Ray images for classification," *Mendeley Data*, vol. 2, no. 2, 2018.
[24] T. Rahman, A. Khandakar, and M. E. H. Chowdhury. (October 2020). Tuberculosis (TB) chest X-ray database. IEEE Dataport. [Online]. Available: https://dx.doi.org/10.21227/mps8-kb56