A Role of Machine Learning Algorithms for Lung Disease Prediction and Analysis

Vishal Borate[1], Dr. Alpana Adsul[2], Palak Purohit[3],
Rucha Sambare[4], Samiksha Yadav[5], Arya Zunjarrao[6]
Assistant Professor, Department of Computer Engineering[1]
Associate Professor, Department of Computer Engineering[2]
Students, Department of Computer Engineering[3,4,5,6]
Dr. D. Y. Patil College of Engineering & Innovation Talegaon, Pune, India

Abstract*: Lung diseases are a notable in global health concern, requiring early diagnosis for better recovery and survival rates. Deep learning strategies, especially CNNs, have shown great promise in self learning lung disease diagnosis from medical images like chest X-rays. Ensemble learning methods using pretrained networks such as VGG16, InceptionV3, and MobileNetV2 have achieved up to 94% accuracy in identifying conditions like COVID-19, pneumonia, and lung opacity. Lightweight CNN models also performed well, with accuracy up to 89.89%. Traditional machine learning algorithms, including Random Forest and Logistic Regression, yielded accuracy rates between 88% and 90%. A hybrid deep learning approach, combining CNN based feature extraction with classifiers like AdaBoost, SVM, and Random Forest, improved classification accuracy by 3.1% and reduced computational complexity by 16.91%. This hybrid method highlights the main feature for integrating deep learning with traditional classifiers to enhance lung disease detection efficiency*

Keywords: Deep learning, Convolutional neural networks (CNNs), Semi-Supervised models, Chest CT images, Classification, Accuracy

I. INTRODUCTION

Among the leading causes of death and disability globally are lung disorders, sometimes known as respiratory infections. The most prevalent of them are COVID-19, pneumonia, and tuberculosis. A recent outbreak of COVID-19 has also been reported. The Forum of International Respiratory Societies' study on the Global Impact of Respiratory Disease states that 1.4 million people die from tuberculosis (TB) annually, affecting over 10 million people. Pneumonia continues to be the biggest cause of death for children under five years old and is another serious killer that claims millions of lives each year. The COVID-19 pandemic is one of the worst in history, having claimed approximately 3.51 million lives worldwide when it first appeared in December 2019. This emphasises how serious lung illnesses are to the world's health.

Respiratory diseases account for five of the top 30 causes of death, emphasizing the urgent need for effective prevention, management, and treatment strategies. Early diagnosis is crucial for improving recovery times and survival rates. Traditionally, chest X-rays and computed tomography (CT) scans have been the primary methods for detecting lung diseases. Chest X-rays are affordable, easy to use, widely accessible, and quicker to perform than CT scans, providing a wealth of information about a patient's condition. However, interpreting these images can be challenging; even skilled radiologists may struggle to differentiate between similar lesions or spot subtle nodules. Additionally, the manual screening process for lung diseases is very slow, need large amount of labour and prone to inconsistencies in interpretation, both among different radiologists and even for the same radiologist at different times.

The shortage of radiologists, coupled with an increasing number of lung disease cases—especially during the COVID-19 pandemic—can lead to delays in diagnosis and treatment. Therefore, there is an urgent need for a robust computer-aided diagnosis (CAD) system to streamline the diagnostic process and make it easier for healthcare professionals. This necessity motivated us to develop a solution that supports radiologists by reducing their task and helping untrained physicians make accurate and timely diagnostic decisions.

Machine learning (ML) is a branch of artificial intelligence (AI) that works on creating statistical models and algorithms that let computers use data to learn from and predict the future. These are some essential machine learning points.

Machine learning is revolutionizing various fields by enabling more efficient data analysis, improving decision-making processes, and creating innovative solutions to complex problems. As the field continues to evolve, the integration of ML into everyday applications is expected to expand, driving advancements in technology and society.

Machine learning is widely used in healthcare, leveraging data mining to reveal hidden patterns in large datasets that aid clinical diagnosis. By analyzing data systematically, healthcare systems can pinpoint inefficiencies and identify best practices to improve patient care and reduce costs. A key focus area is lung disease detection, leading many researchers to develop advanced medical decision support systems to enhance physicians' diagnostic and treatment abilities.

Therefore, we reviewed various research papers, surveys and theories concerning machine learning algorithms used to predict lung diseases. So, we identified several common machine learning and deep learning algorithms employed for prediction, including Bagging, Logistic Regression, Random Forest, Support Vector Machines (SVM), and Naive Bayes. These algorithms contribute to effective outcomes in lung disease prediction.

II. RELATED WORK

An In paper [1] Ishan Sen,et. al. Focuses on predicting lung diseases using machine learning algorithms, addressing conditions like asthma, COPD, bronchitis, emphysema, and lung cancer. The dataset comprises 323 instances and 19 attributes, representing patients with a range of lung diseases and associated symptoms. Five machine learning algorithms were employed: Bagging, Logistic Regression, Random Forest, Logistic Model Tree, and Bayesian Networks, with model training conducted using K-Fold Cross Validation. The algorithms achieved precision rates of 88.00% (Bagging), 88.92% (Logistic Regression), 90.15% (Random Forest), 89.23% (Logistic Model Tree), and 83.69% (Bayesian Networks), with Random Forest being the model that execute the finest. This research demonstrates machine learning's potential in early detection and prediction of lung diseases, which could aid in proactive healthcare measures.

In paper [2] Muhab Hariri et al. Presents a lightweight convolutional neural network (CNN) for diagnosing lung diseases, including COVID-19 and pneumonia, using chest X-ray images. The model achieved 89.89% accuracy on a public dataset, outperforming the Efficient-Net B2 model at 85.7%. While both models effectively identified COVID-19, they struggled to distinguish between viral and bacterial pneumonia due to limited viral pneumonia samples. The model's low computational requirements make it suitable for resource-limited medical settings and IoT applications. Future work may focus on enhancing image quality and addressing data imbalances to improve viral pneumonia detection.

In paper [3] Mona Hmoud Al-Sheikh et. al. Suggests an self-operating system that uses a customized (CNN), two pre-trained deep learning models (Alex Net, VGG16Net), along with a novel image enhancement algorithm, to detect multiple lung diseases in chest X-ray and CT scans. The system consists of two key steps: image pre-processing, where a k-symbol Lerch transcendent function is applied to enhance images, and deep learning classification. Tested on public datasets, the method achieved high accuracy (98.60% for X-Ray, 98.80% for CT), demonstrating the effectiveness of the image enhancement and classification models.

In paper [4]Patel M et. al. Develops a robust deep learning-based system for identifying lung diseases, including COVID-19, pneumonia, lung opacity, and normal states, from chest X-ray images. Using a Kaggle dataset, the images are preprocessed for contrast enhancement and noise removal, with near-miss resampling addressing dataset imbalance. Ensemble learning techniques with VGG16, InceptionV3, and MobileNetV2 are used, achieving 94% accuracy with a three-level ensemble method. The system includes a web interface for remote access. This research demonstrates the capability of ensemble deep learning for automated, profitable lung disease detection, providing valuable decision support for healthcare, especially in resource-limited settings.

In paper [5]Nivethitha. A et. al. Discusses a disease forecast system that employ machine learning to predict illnesses based on user-entered symptoms. The system not only predicts diseases but also offers health advice, leveraging machine learning models like Random Forest, Naïve Bayes, Decision Tree, and K-Nearest Neighbor (KNN). The prediction process is facilitated through Python and Tkinter for the user interface, using a dataset from hospitals for

training. The system evaluates results based on metrics like accuracy, sensitivity, specificity, positive predictive value, and negative predictive value, achieving an overall accuracy of around 95%. It provides a user-friendly interface and stores data for future improvements.

In paper [6]Sadiya, A. et. al. Explores the utilization of machine learning approach for the differential determine of Tuberculosis and Pneumonia, diseases with similar symptoms. The study develops a classification model using Naïve Bayes, Decision Tree, and Random Forest algorithms, with the highest accuracy improvement seen with Naïve Bayes and overall best performance from Random Forest. The results show a slight performance increase with discretization techniques. Future work aims to extend the model to other diseases with similar symptoms and improve accuracy through additional preprocessing and discretization methods.

In paper [7] Gunasinghe, A.D. et. al. discusses using machine learning and deep learning techniques such as CNN, image processing to develop algorithms for early detection and diagnosis of lung diseases, including asthma, COPD, tuberculosis, pneumothorax, and lung cancer. By analyzing patient data and chest X-ray images, the project aims to create a binary classification model that assists doctors in making timely diagnostic decisions to improve patient outcomes. In future work, this plans to train the model with larger datasets and adjust parameters to enhance processing speed. Additional performance metrics will be evaluated, and pre-trained models may be experimented with to further improve accuracy.

In paper [8] Kadir, T. et. al. Discusses the employment of CNNs and deep learning for nodule classification and lung cancer prediction from CT imaging, achieving high AUC performance. It emphasizes the need to define how CADx outputs should be integrated into clinical decision-making, addressing questions about risk assessment, algorithm incorporation into guidelines, and comparisons with radiologists. It also emphasizes the importance of considering the characteristics of training and validation datasets, such as smoking history or malignancy.

In paper [9] M.J. Wildman et. al. Develops COPD and Asthma Physiology Score (CAPS) was developed using logistic regression analysis on 8,527 individuals with obstructive airway disease (mean age 65.9 years, hospital mortality 35.5%). With an AUC of 0.718, CAPS, which measures eight different variables (heart rate, MAP, pH, salt, urea, creatinine, albumin, and WBC count), demonstrated fair discrimination. This performance exceeded SAPS II, APACHE II, and APACHE III ratings in 7,957 patients that were validated.

In paper [10]Stephen A. Deppen et. al. Reviews the limitations of existing lung cancer prediction models for patients being evaluated for surgery and the need for models that address this specific population. By developing the TREAT model, a clinical prediction model for lung cancer. Authors used logistic regression to develop TREAT (Thoracic Research Evaluation and Treatment) model reaching upto 87% accuracy which is measured through AUC (Area under the ROC curve). Also comparing the developed model to existing mayo clinic model which gives 80% of accuracy.

In paper [11]Francesco Ciompi et. al. Introduces deep learning system using multi-stream, multi-scale convolutional neural networks to automatically classify pulmonary nodules in CT scans and to process multiple 2D views of nodules across different scales, it also compares against classical machine learning methods like Support Vector Machines (SVMs) using intensity and unsupervised features. Gaining overall accuracy of 79.5% and F-measure (harmonic mean of precision and recall) per class ranged from 43.4% for spiculated nodules to 85.7% for calcified nodules. Although facing an issue of data imbalance, talking about other such existing models like PanCan Model, Lung-RADS Guidelines, KNN (k-nearest neighbors), SIFT (Scale-Invariant Feature Transform).

In paper [12] Sudipta Mukhopadhyay et. al. Converses about segmentation framework for pulmonary nodules in lung CT images, focusing on different types of nodules such as solid, part-solid, and non-solid aiming to improve the segmentation accuracy for each type, considering internal texture and external attachment to pleura or vessels. iterative morphological filtering, vasculature pruning technique, 3D region growing with fuzzy connectivity, Ellipsoid Approximation, Selective Enhancement Filtering addressing the challenges of segmenting different types of pulmonary nodules and introduces a robust framework that adapts based on the nodule type. She also discusses and borrows some ideas from existing models such as Kuhnigk, Moltz, Kubota, Dehmeshki. The proposed method achieved the highest accuracy for both solid and non-solid nodules compared to existing techniques, accuracy of 99% for solid and part-solid nodules and 98% for non-solid nodule.

In paper [13] Annette McWilliams et. al. Outlines the methods used for data collection and validation in the (ICNARC) Case Mix Programme (CMP), which assesses outcomes for adult critical care admissions. The CMP ensures data

accuracy through several steps, such as standardized dataset specifications, training courses, and both local and central data validation. It applies established criteria from the Directory of Clinical Databases to assess data quality. several techniques used are data specification, validation, Evaluation Against Criteria that is CMP data was evaluated using criteria from the Directory of Clinical Databases, scoring a mean quality level of 3.4 out of 4. The results showed the CMPD contained validated data on over 129,000 admissions, with a variety of clinical metrics. The database is highly reliable and comparable to other UK critical care datasets.

In paper [14] Jason L. et. al. Outlines the methodology and statistical analysis plan for creating and assessing clinical prediction models for pulmonary nodules. Where the limitations of existing models are discussed.

In paper [15] Jingsi Dong et. al Develops a logistic regression diagnostic model to discriminate between benign and malignant solitary pulmonary nodules (SPNs). The model was constructed using clinical, biomarker, and radiological data from a training set of 1,679 patients (77.2% with malignant SPNs). Independent variables included factors such as age, smoking history, family cancer history, nodule diameter, and CT characteristics (e.g., spiculation, lobulation, calcification). Techniques used include logistic regression for multivariable analysis and ROC (Receiver Operating Characteristic) curves to evaluate the model's performance in both training and test datasets. Additionally, the study compared this model with the Swensen model and the Li model, both of which also use logistic regression but with different independent factors and achieved high accuracy.

In paper [16] Gary S Collins et. al. Discusses the development of the TRIPOD which stands for Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis Initiative, which aims to improve the quality and transparency of reporting in studies related to prediction models. These models help healthcare providers estimate the risk of a disease (diagnostic models) or predict future events (prognostic models). Despite their importance, prediction model studies often suffer from poor reporting, making it difficult to assess their bias and usefulness. The goal of TRIPOD is to enhance the standard and clarity of reporting in prediction model research, making it easier for readers and editors to evaluate the studies.

In paper [17] Heber MacMahon et. al. Describes the creation of the Lung Image Database Consortium that is LIDC and Image Database Resource Initiative (IDRI), which developed a publicly available repository of computed tomography (CT) scans for lung nodule detection and assessment. This initiative involved collaboration among seven academic centres and eight medical imaging companies, resulting in a robust database of 1,018 clinical thoracic CT cases. Each case includes images and an XML file detailing a two-phase annotation process carried out by four thoracic radiologists are included with each case. In the initial phase, radiologists independently categorized lesions as "nodule greater than or equal to 3 mm," "nodule less than 3mm," or "non-nodule." In the second phase, they analysis their own marks alongside those of their peers to reach a final opinion. The database contains 7,371 lesions marked as nodules, with 2,669 classified as "nodule greater than or equal to 3 mm." This database aims to enhance medical imaging research related to lung nodules.

In paper [18] Anastasia Chalkidou et. al. Talks through the usage of texture features in PET and CT imaging for cancer prognosis using benjamini-hochberg correction, optimum cut-off approach, kaplan-meier survival analysis concluding that the association between patient survival and texture features on PET or CT scans is not well supported by the available data.

In paper [19] Daniel Hammack Talks about using deep learning models like 17 3D convolutional neural network along with nodule detection to identify the abnormal regions within the lung scans. The core of the system involves 17 different 3D CNN models, organized into two ensembles, which were combined for better performance, the models used Keras and Theano frameworks, Leaky ReLU activation functions and softplus regression layers, batch normalization, 3D convolution, and global max pooling layers, Extremely Randomized Trees (ERT) were used in a final classifier, which worked with nodule features predicted by the CNN models. The system also applied test-time augmentation. However the model was based on a limited dataset and a heavy computation resources were used. The challenges involve dealing with high resolution medical images.

In paper [20] Hugo J.W.L. et. al. Presents a radiomics approach to decode tumor phenotypes using non-invasive imaging, primarily focusing on lung and head-and-neck cancers, authors here used a quantitative radiomics method to extract 440 image features from CT scans. These features quantify tumor intensity, shape, texture, and wavelet decomposition where a multivariate Cox proportional hazards regression model was built using the top four radiomic

features for prognosis. Also taking us through existing technologies like RECIST and WHO measure tumor response to therapy using one- or two-dimensional size descriptors and TNM staging focuses on tumor resect ability. The developed model accomplished 0.65 for lung cancer data and 0.69 for head-and-neck cancer data.

In paper [21] Patrick Royston et. al. Reviews the use of Multiple Imputation (MI) in 16 studies, highlighting variability in techniques and handling of missing data. Most studies used five or ten imputations, with no consensus on software. Rubin's rules were commonly used to combine estimates after MI, though some studies didn't specify how this was done. Reported estimates included regression coefficients, hazard ratios, and confidence intervals from models like Cox regression, Poisson regression, and Weibull models. Some studies reported model performance but without detailing post-imputation calculations.

In paper [22] Warren S. Browner et. al. Concludes, these findings indicate that the random assignment of heavy smokers to a smoking cessation program results in a significant reduction in the rate of pulmonary function decline, specifically a decrease of approximately 16%, corresponding to an improvement of over 10 ml/year in FEV1.[38][39] Although we believe these results can be extrapolated to lighter smokers and potentially to female populations, validating this effect through randomized designs remains complex. Importantly, our data suggest that even among heavy and chronic smokers, cessation can yield beneficial effects, emphasizing that it is not too late to quit.[37][40]

In paper [23] M.E.J. Callister et. al. Takes us through risk assessment of malignancy in pulmonary nodules based on clinical and radiological factors, imaging techniques, follow-up protocols, and guidelines for surgical and non-surgical biopsy.[36] Introduction of malignancy prediction calculators and improved algorithms to refine the risk assessment and management of pulmonary nodules. As well as provides detailed guidelines on the management of sub-solid nodules (SSNs).[35][41]

In paper [24] Salma Sultana et. al. Studies three pre-trained transfer learning models—EfficientNetB0, DenseNet169, and DenseNet201—were implemented to classify lung conditions, including COVID-19, Pneumonia, Tuberculosis, and healthy lungs.[34] The dataset, consisting of 6,340 images, was split into training, testing, and validation sets in an 80:15:5 ratio. Image preprocessing techniques such as resizing, filtering, and augmentation were applied. Model performance was evaluated using confusion matrices and metrics like accuracy, precision, recall, F1-score, and rates such as TPR, TNR, FPR, and FNR to determine the optimal classification model.[33] Young scientists can also learn from this paper how to develop CNN models that can be used to identify diseases early on in the process of using medical images.[32]

In paper [25] Hole, D.J., et. al. Deliberates a follow-up study that found 2,545 men and 1,894 women died, with increasing mortality risk linked to lower FEV1 after adjusting for age, smoking, blood pressure, cholesterol, BMI, and social class.[26] Relative hazard ratios for all-cause mortality in the lowest FEV1 quintile were 1.92 for men and 1.89 for women.[27] Reduced FEV1 was also significantly associated with higher risk of death from ischemic heart disease, lung cancer, and stroke.[28] For lifelong nonsmokers, reduced FEV1 was linked to all causes of death except cancer.[29] Impaired lung function is a strong indicator of mortality risk, highlighting the importance of FEV1 in health assessments, especially for smokers.[30] This study serves as a foundation for exploring the potential of using machine learning algorithms to improve the detection and prediction of lung diseases, aiming for more accurate and reliable outcomes.[31]

III. PROPOSED ARCHITECTURE

In a lung disease prediction system based on GANs and CNNs, the process begins with acquiring high-resolution medical images, such as CT scans or X-rays, from datasets like ChestX-ray14 or LUNA16. These images undergo preprocessing, which includes resizing, normalization, and noise reduction to prepare them for analysis. A Generative Adversarial Network (GAN) is then used for data augmentation. The generator (G) learns to create synthetic lung images from random noise vectors, improving over time by minimizing the error in fooling the discriminator (D), which distinguishes between real and generated images. Once trained, the augmented synthetic images, alongside real ones, are passed into a CNN. The CNN extracts relevant features through convolution and pooling layers, learning patterns in the data to classify lung images as healthy or diseased (e.g., pneumonia, tuberculosis). Finally, model evaluation using metrics such as accuracy, precision, recall, and F1-score ensures that the system is fine-tuned for

accurate disease prediction. This integration of GAN-generated data with CNN-based classification enhances the overall predictive performance of the system.
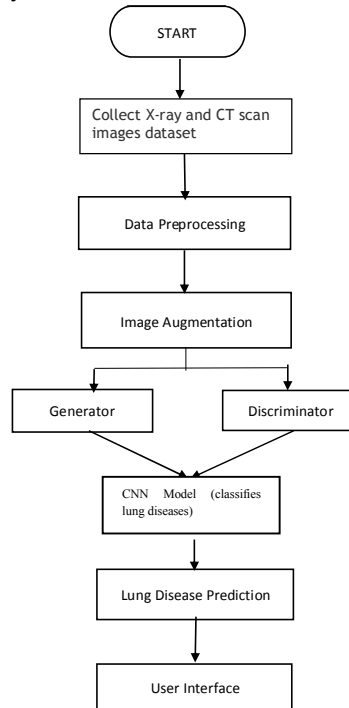
IV. EXPERIMENTAL SETUP

To conduct the experiment, specific minimum hardware requirements must be met, including a solid-state drive (SSD) with a capacity of 512 GB for storage, CPU that is central processing unit accompanied by a graphics processing unit (GPU), and a minimum memory (RAM) of 16 GB, with 32 GB recommended for optimal performance. Additionally, an adequate cooling system is necessary to maintain optimal operating temperatures. The software requirements consist of a Windows operating system, a development environment utilizing Jupyter Notebook and Python version 3.7, and essential machine learning libraries and frameworks such as TensorFlow, scikit-learn, pandas, and OpenCV. For model deployment, frameworks like Flask or Django should be employed to facilitate application deployment.

IV. RESULT AND ANALYSIS OF ALGORITHMS

TABLE I Algorithm/Author	Precision (%)	Recall (Sensitivity %)	Accuracy (%)
Kuhnigk et al.	48	39	99
Moltz et al.	58	48	99
Kubota et al.	50	56	97
Proposed Model	62	58	99

Accuracy: It is the percentage of accurate predictions true positives and true negatives that are made relative to all of the forecasts that were made. It indicates how frequently the model is accurate overall.

Accuracy= True Positives + True Negatives

Total Predictions

Precision: It measures the accuracy of positive predictions how many of the predicted positives are actually correct.

Precision= True Positives + False Positives

True Positives

Recall: It gauges how well a model can detect every positive case and is also referred to as sensitivity or true positive rate. It provides an answer to the following query: how many real benefits did the model accurately identify.

$$\text{Recall} = \frac{\text{True Positives} + \text{False Negatives}}{\text{True Positives}}$$

Kuhnigk[12], this method achieves a high accuracy of 99%, indicating that it correctly identifies the majority of nodules. However, its precision and recall are relatively low, with precision at 48% and recall at 39%. This suggests that while the model is accurate overall, it struggles with correctly identifying true positive cases, leading to a higher rate of false positives and lower sensitivity.

Similar to Kuhnigk et al.[12], Moltz[18] model also boasts an accuracy of 99%. However, it has improved precision at 58% and recall at 48%. This indicates that it performs slightly better in correctly identifying positive cases compared to Kuhnigk[12] et al., but still suffers from challenges in sensitivity.

Kubota[19] method has a slightly lower accuracy of 97% compared to the previous two models. Its precision stands at 50%, while recall is better at 56%. This means it is more effective at identifying true positives than the prior methods, though its overall accuracy is slightly lower.

The Model proposed by Sudipta Mukhopadhyay[12] achieves the same high accuracy of 99% as the previous methods but excels in both precision (62%) and recall (58%). This indicates that it is more effective at identifying true positives while maintaining a high overall accuracy.

V. CONCLUSION

Thus, we conclude from above that machine learning methods can effectively classify lung diseases. Among the various algorithms, Random Forest performs better than Logistic Regression, Bagging, Logistic Model Trees (LMT), and Naive Bayes, with accuracy rates of 88.00%, 88.92%, 90.15%, 89.23%, and 83.69%, respectively. Nonetheless, the challenge of acquiring real-time data persists. Notably, dataset includes numerous attributes, which is uncommon in online sources. Furthermore, several research papers indicate that future work may focus on utilizing advanced machine learning algorithms like Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), and Artificial Neural Networks (ANN) to achieve even more precise and accurate results.

By reviewing various studies, we aim to propose a method for predicting lung disease that addresses the challenge of data scarcity through the use of diverse algorithms and techniques. We expect that this method will enhance the efficiency and accuracy of the model in identifying lung diseases, ultimately improving patient outcomes and long-term survival rates).

REFERENCES

[1]. Sen, Ishan & Hossain, Md & Shakib, Faisal & Imran, Asaduzzaman & Faisal, Faiz. (2020). In Depth Analysis of Lung Disease Prediction Using Machine Learning Algorithms. 10.1007/978-981-15-6318-8_18.

[2]. Hariri, Muhab & Avşar, Ercan. (2023). COVID-19 and pneumonia diagnosis from chest X-ray images using convolutional neural networks. Network Modeling Analysis in Health Informatics and Bioinformatics. 12. 10.1007/s13721-023-00413-6.

[3]. Ahmad Sami Al-Shamayleh, Hamid A Jalab, Mona Hmoud Al-Sheikh, Omran Al Dandan, Rabha W Ibrahim. Multi-class deep learning architecture for classifying lung diseases from chest X-Ray and CT images. Scientific Reports. 13 (2023,11).

[4]. Patel M, Shah M (2024) Deep Ensemble Learning Model for Diagnosis of Lung Diseases from Chest X -Ray Images . Indian Journal of Science and Technology 17(8): 702-712. https://doi.org/10.17485/IJST/v17i8.3151.

[5]. Nivethitha. A, Pramoth Krishnan. T, Narendran. G. Smart Disease Prediction Using Machine Learning. International Journal of Innovative Science and Research Technology ISSN No:-2456-2165.

[6]. S. Modi, Y. K. Mali, V. Borate, A. Khadke, S. Mane and G. Patil, "Skin Impedance Technique to Detect Hand-Glove Rupture," 2023 OITS International Conference on Information Technology (OCIT), Raipur, India, 2023, pp. 309-313, doi: 10.1109/OCIT59427.2023.10430992.

[7]. Y. Mali, M. E. Pawar, A. More, S. Shinde, V. Borate and R. Shirbhate, "Improved Pin Entry Method to Prevent Shoulder Surfing Attacks," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-6, doi: 10.1109/ICCCNT56998.2023.10306875

[8]. A Chaudhari et al., "Cyber Security Challenges in Social Meta-verse and Mitigation Techniques," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-7, doi: 10.1109/MITADTSoCiCon60330.2024.10575295.

[9]. Y. K. Mali and A. Mohanpurkar, "Advanced pin entry method by resisting shoulder surfing attacks," 2015 International Conference on Information Processing (ICIP), Pune, India, 2015, pp. 37-42, doi: 10.1109/INFOP.2015.7489347 '.

[10]. Y. K. Mali, S. A. Darekar, S. Sopal, M. Kale, V. Kshatriya and A. Palaskar, "Fault Detection of Underwater Cables by Using Robotic Operating System," 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, 2023, pp. 1-6, doi: 10.1109/ICCST59048.2023.10474270.

[11]. A Chaudhari, S. Dargad, Y. K. Mali, P. S. Dhend, V. A. Hande and S. S. Bhilare, "A Technique for Maintaining Attribute-based Privacy Implementing Blockchain and Machine Learning," 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, 2023, pp. 1-4, doi: 10.1109/ICCST59048.2023.10530511.

[12]. Y. K. Mali, S. Dargad, A. Dixit, N. Tiwari, S. Narkhede and A. Chaudhari, "The Utilization of Block-chain Innovation to Confirm KYC Records," 2023 IEEE International Carnahan Conference on Security Technology (ICCST), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICCST59048.2023.10530513.

[13]. V. Borate, Y. Mali, V. Suryawanshi, S. Singh, V. Dhoke and A. Kulkarni, "IoT Based Self Alert Generating Coal Miner Safety Helmets," 2023 International Conference on Computational Intelligence, Networks and Security (ICCINS), Mylavaram, India, 2023, pp. 01-04, doi: 10.1109/ICCINS58907.2023.10450044.

[14]. M. Dangore, A. S. R, A. GhanashyamChendke, R. Shirbhate, Y. K. Mali and V. Kisan Borate, "Multi-class Investigation of Acute Lymphoblastic Leukemia using Optimized Deep Convolutional Neural Network on Blood Smear Images," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-6, doi: 10.1109/MITADTSoCiCon60330.2024.10575245.

[15]. N. P. Mankar, P. E. Sakunde, S. Zurange, A. Date, V. Borate and Y. K. Mali, "Comparative Evaluation of Machine Learning Models for Malicious URL Detection," 2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon), Pune, India, 2024, pp. 1-7, doi: 10.1109/MITADTSoCiCon60330.2024.10575452.

[16]. M. D. Karajgar et al., "Comparison of Machine Learning Models for Identifying Malicious URLs," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/ICITEICS61368.2024.10625423.

[17]. J. Pawar, A. A. Bhosle, P. Gupta, H. Mehta Shiyal, V. K. Borate and Y. K. Mali, "Analyzing Acute Lymphoblastic Leukemia Across Multiple Classes Using an Enhanced Deep Convolutional Neural Network on Blood Smear," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/ICITEICS61368.2024.10624915.

[18]. D. R. Naik, V. D. Ghonge, S. M. Thube, A. Khadke, Y. K. Mali and V. K. Borate, "Software-Defined-Storage Performance Testing Using Mininet," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-5, doi: 10.1109/ICITEICS61368.2024.10625153.

[19]. A O. Vaidya, M. Dangore, V. K. Borate, N. Raut, Y. K. Mali and A. Chaudhari, "Deep Fake Detection for Preventing Audio and Video Frauds Using Advanced Deep Learning Techniques," 2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Kothamangalam, Kerala, India, 2024, pp. 1-6, doi: 10.1109/RAICS61201.2024.10689785.

[20]. Bhongade, A., Dargad, S., Dixit, A., Mali, Y.K., Kumari, B., Shende, A. (2024). Cyber Threats in Social Metaverse and Mitigation Techniques. In: Somani, A.K., Mundra, A., Gupta, R.K., Bhattacharya, S., Mazumdar, A.P. (eds) Smart Systems: Innovations in Computing. SSIC 2023. Smart Innovation, Systems and Technologies, vol 392. Springer, Singapore. https://doi.org/10.1007/978-981-97-3690-4_34.

[21]. Sawardekar, S., Mulla, R., Sonawane, S., Shinde, A., Borate, V., Mali, Y.K. (2025). Application of Modern Tools in Web 3.0 and Blockchain to Innovate Healthcare System. In: Rawat, S., Kumar, A., Raman, A., Kumar, S., Pathak, P. (eds) Proceedings of Third International Conference on Computational Electronics for Wireless Communications. ICCWC 2023. Lecture Notes in Networks and Systems, vol 962. Springer, Singapore. https://doi.org/10.1007/978-981-97-1946-4_2

[22]. Mali, Y., &Chapte, V. (2014). Grid based authentication system, International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 10, October 2014 pg. 93-99, 2(10).

[23]. Yogesh Mali, NilaySawant, "Smart Helmet for Coal Mining", International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 3, Issue 1, February 2023, DOI: 10.48175/IJARSCT-8064

[24]. Pranav Lonari, SudarshanJagdale, Shraddha Khandre, PiyushTakale, Prof Yogesh Mali, "Crime Awareness and Registration System ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 3, pp.287-298, May-June-2021.

[25]. Jyoti Pathak, Neha Sakore, Rakesh Kapare , Amey Kulkarni, Prof.Yogesh Mali, "Mobile Rescue Robot", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4, Issue 8, pp.10-12, September-October-2019.

[26]. DevanshDhote , Piyush Rai , Sunil Deshmukh, AdarshJaiswal, Prof.Yogesh Mali, "A Survey : Analysis and Estimation of Share Market Scenario ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4, Issue 8, pp.77-80, September-October-2019.

[27]. RajatAsreddy, AvinashShingade, NirajVyavhare, Arjun Rokde, Yogesh Mali, "A Survey on Secured Data Transmission Using RSA Algorithm and Steganography", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 4, Issue 8, pp.159-162, September-October-2019

[28]. ShivaniChougule, ShubhamBhosale, VrushaliBorle, VaishnaviChaugule, Prof.Yogesh Mali, "Emotion Recognition Based Personal Entertainment Robot Using ML & IP", International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, Volume 5, Issue 8, pp.73-75, November-December-2020.

[29]. Amit Lokre, SangramThorat, PranaliPatil, ChetanGadekar, Yogesh Mali, " Fake Image and Document Detection using Machine Learning", International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X, Volume 5, Issue 8, pp.104-109, November-December-2020.

[30]. RiteshHajare, RohitHodage, Om Wangwad, Yogesh Mali, FarazBagwan, "Data Security in Cloud", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 3, pp.240-245, May-June-2021

[31]. Yogesh Mali and TejalUpadhyay, "Fraud Detection in Online Content Mining Relies on the Random Forest Algorithm", SWB, vol. 1, no. 3, pp. 13–20, Jul. 2023, doi: 10.61925/SWB.2023.1302.

[32]. Mali, Y. K., Rathod, V. U., Borate, V. K., Chaudhari, A., &Waykole, T. (2023, June). Enhanced Pin Entry Mechanism for ATM Machine by Defending Shoulder Surfing Attacks. In International Conference on Recent Developments in Cyber Security (pp. 515-529). Singapore: Springer Nature Singapore.

[33]. Mali, Y. K., Rathod, V., Dargad, S., &Deshmukh, J. Y. (2024). Leveraging Web 3.0 to Develop Play-to-Earn Apps in Healthcare using Blockchain. In Computational Intelligence and Blockchain in Biomedical and Health Informatics (pp. 243-257). CRC Press.

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-19962

ISSN
2581-9429
IJARSCT

433

[34]. Deshmukh, J. Y., Rathod, V. U., Mali, Y. K., & Sable, R. (2024). and Classification. Data-Centric Artificial Intelligence for Multidisciplinary Applications, 114.

[35]. Sawant, M. M., Nagargoje, Y., Bora, D., Shelke, S., & Borate, V. (2013). Keystroke Dynamics. International Journal of Advanced Research in Computer and Communication Engineering, 2(10), 4018-4020.

[36]. Ingale, N., Jha, T. A., Dixit, R., & Borate, V. K. (2021). College enquiry Chatbot using rasa. College Enquiry ChatBot Using RASA.

[37]. Patil, Yash, MihirPaun, Deep Paun, Karunesh Singh, and Vishal Kisan Borate. "Virtual painting with OpenCV using Python." International Journal of Scientific Research in Science and Technology 5, no. 8 (2020): 189-194.

[38]. Borate, V. K., &Giri, S. (2015, January). XML duplicate detection with improved network pruning algorithm. In 2015 International Conference on Pervasive Computing (ICPC) (pp. 1-5). IEEE.

[39]. Shabina Modi, Sunita Mane, SakshiMahadik, RutujaKadam, RutujaJambhale, SampadaMahadik, &Yogesh Mali. (2024). Automated Attendance Monitoring System for Cattle through CCTV. Revista Electronica De Veterinaria, 25(1), 1025 -1034. https://doi.org/10.69980/redvet.v25i1.724

[40]. Sawardekar, S., Mulla, R., Sonawane, S., Shinde, A., Borate, V., & Mali, Y. K. (2023, December). Application of Modern Tools in Web 3.0 and Blockchain to Innovate Healthcare System. In International Conference on Computational Electronics for Wireless Communications (ICCWC) (pp. 1-12). Singapore: Springer Nature Singapore.

[41]. A More, S. R. Shinde, P. M. Patil, D. S. Kane, Y. K. Mali and V. K. Borate, "Advancements in Early Detection of Lung Cancer using YOLOv7," 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2024, pp. 1739-1746, doi: 10.1109/ICOSEC61587.2024.10722534.