Grenze International Journal of Engineering and Technology, January Issue

AI Methods for Lung Disease Diagnosis: A Comprehensive Survey

Prof. Roshan Bhanuse[1], Samiksha V. Sawant[2], Sejal Mahadule[3], Rutuja Kute[4], Tanushree Darbeshwar[5] and Khushi Latey[6]
[1-6]Dept. of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur 441110

***Abstract*— Continuous exposure from breathing in coal dust, and other hazardous airborne particulates generally has a way of causing many serious lung diseases such as pneumoconiosis and chronic obstructive pulmonary disorder (COPD) in coal miners. Thus, identifying risks as early as we can is essential, so that remedial or preventative steps protecting at risk employees can be taken. To find out the risk of these diseases in coal miners, this paper reviews 75 studies that explore various predictive methods.**

The studies reviewed use a range of research designs, from the conventional statistical models to more advanced machine learning algorithms, including neural networks, logistic regression, support vector machines, and random forests. The target diseases are primarily COPD and pneumoconiosis, and the data sources are clinical test data, environmental exposure records, and biochemical indicators in blood. The amount of time a miner faces riskier environments, specific characteristics of their environment (like angle of the coal seams), the measurements they collected on their lung activity (like flow-volume loops and vital capacity), and biological indicators (like cortisol levels, and plasma antioxidant/reductant activity) would need to be addressed to make disease predictions. The models that were reviewed were 70% to 91% accurate. However, many of the models specified were developed using the original datasets, and consequently are not applicable to a larger audience. Additionally, the review highlights major deficiencies in moving forward by addressing the limitations of comparing multiple data pertaining to diverse components of their health and working contexts. There is a need for valid predictive models that are potentially generalizable to larger groups [11],[12]. By synthesizing clinical data with both environmental exposure and biochemical data, lung health may be preserved for coal miners; and the potential to improve the detection of early disease is promising.

***Keywords*— lung health and coal miners, biomarkers and COPD, pneumoconiosis, predictive modeling, early detection of lung disease, machine learning and medicine, health of coal miner.**

I. INTRODUCTION

The health of a economy and developed community's working population is essential. Not only is it advantageous for all those involved to protect their workers from work hazards, but a more compelling reason is the impact workplace hazards can have on maintaining a productive workforce, especially when many workers are exposed to various physically demanding and hazardous conditions. To say coal miners have the highest impact on their respiratory health from high-hazard jobs is an understatement.

Grenze ID: 01.GIJET.12.1.27_2

In addition to the length of exposure to coal dust and other air-borne particulate matter, both the chronic obstructive pulmonary disease (COPD) and pneumoconiosis, also referred to as black lung disease or silicosis, are prevalent diseases found among coal miners and their exposure to coal dust as well as speculated particulate lung disease. [7][10] Thankfully, we now have dust mitigation plans and better ventilation systems. Alas, the results of ongoing research indicate there are still unforeseen residuals of these diseases that perpetuate in the some mining towns around the world.

Frequent inhalation of small fibrous dust particles deep into the lungs is the cause of many respiratory illnesses. Over the years, a tethered supply of lung damage, which is irreversible. This can significantly lose a person's capacity to breathe, work, and live their life normally. Coal is a significant contributor to both the development of many of the occupational illnesses. Long-standing history and near-depth research—many of which have been conducted in the Donbass—have consistently shown this. These studies have also presented widespread knowledge about how dust is formed and distributed underground, as well as what measures may help to control it.

A. Rationale Behind this Survey

This evaluation is well-timed because, despite a number of longer-term problems, coal mining operations have evolved, primarily through new equipment and the incorporation of new mining method options. On the other hand, the risk of respiratory disease still exists.

Moreover, global illnesses such as the COVID-19 pandemic raised concerns of worker safety, and safety of workers used to high levels of exposure, specifically because of air borne dust. Thus, this study synthesizes recent studies and aims to evaluate the efficacy of current dust management practices, legislation, and health surveillance for protecting miners lungs.

B. Contribution of the Survey

This paper thoroughly analyzes respiratory diseases as they relate to coal dust and summarizes data that has been collected across more than 60 existing studies and scientific research. This research paper will specifically evaluate the effectiveness based on real-world situations of some of the interventions - such as; dust suppression equipment, Section 431.92, and medical screening - available to deal with respiratory illness. The paper will also consider, apart from respiratory illness, a public health crisis that is frequently overlooked within broader discussions of the mining industry, which is the health impacts for citizens and communities who live next door to the coal mining industry.

C. What Set this Survey Apart

Past assessments appear to have weighed one condition, either pneumoconiosis or COPD, at a time, while this evaluation is simply considering them together and thinking about both as an integrated phenomenon. The last thing this evaluation does is think about the last few decades in terms of addition where natural resources are being more effectively managed and improved, the policies for workers are in line and cohesive, and changes in the environmental policies have not been examined in other research. The intention here is to more effectively describe the variety of hazards that miners and others are exposed to and what needs to be done to protect their health.

Finally, this assessment intends to help the policy makers, researchers and health care professionals determine the proper means to mitigate the coal health hazards of all miners, whether they are underground coal miners or surface coal miners.

II. Background Work

This research looks into the changes made in the recent years regarding the use of deep learning techniques for human activity recognition (HAR), focusing on how modern approaches are designed to capture the spatiotemporal features of sensor data. The applied field of HAR developed from simple static feature engineering followed by the application of standard machine learning techniques like SVM, Decision Trees, and k-NN. These techniques relied on handcrafted features which concentrated on domain knowledge and did not generalize well over studies and datasets.

The discipline has witnessed numerous breakthroughs due to the application of deep learning, and more specifically, CNNs and RNNs. Spatial features of raw sensor data have been successfully extracted by CNNs, while RNNs, and more specifically LSTM networks, have performed well in capturing the temporal aspect of sequences.

This study aims to develop a novel hybrid deep learning architecture composed of CNN and LSTM layers, incorporating techniques to simultaneously optimize for spatial and temporal feature extraction. The model complements the push towards more recent trends,

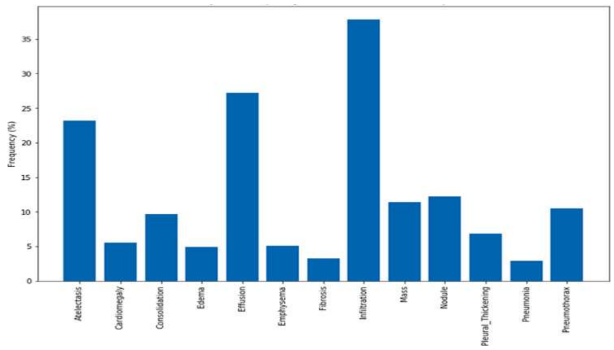

Fig. 1. Adjusted frequency of diseases in patient group on the image dataset.

III. RECOMMENDATION TECHNIQUES

Four primary recommendation strategies have emerged in recent research:

A. Machine learning-Based Strategies, Several supervised ML, method have demonstrated promising results, including

- Support Vector Machine (SVM)
- Random Forest
- Logistic Regression
- Bayesian Networks

B. Deep Learning & CNN-Based Strategies Recommended deep learning model include

- Convolutional Neural Networks (CNN)
- Deep Neural Networks (DNN)
- Generative Adversarial Networks (GANs)

C. Pathology and Biomarker-Based Techniques for occupational lung disease, recommended procedure include

- Lung biopsies (cryobiopsy, surgical, transbronchial)
- Bronchoalveolar lavage (BAL)
- Polarized light microscopy

D. Optimization, Preprocessing & System-Level Strategies effective data augmentation techniques include

- Horizontal flip
- Rotation
- Zoom
- Shear

IV. LITERATURE REVIEW

TABLE I

Paper Name	Technique	What It Does	Pros	Cons	Accuracy	Dataset Source
Deep learning-based detection of silicosis (Aksoy et al.) **[10]**	Transfer Learning (Xception, DenseNet121, etc.)	Detects silicosis in CT images	High accuracy, compares 6 DL models	Requires CT scans, high resources	97.29% (Xception)	Custom CT dataset from silicosis patients, Turkey
Prediction of Silicosis in Iron and Steel Workers (Safwat & Nagy) **[6]**	Ensemble (FFNN + KNN + Fuzzy KNN + DT)	Predicts silicosis risk from health & exposure data	Handles mixed features, high performance	Region-specific (Egypt), questionnaire -based	97.2%	El Gedida Iron Mine, Egypt; 220 workers

Hybrid DL for Lung Diseases (Bharati et al.) **[12]**	VGG + STN + CNN (VDSNet)	Classifies lung diseases from X- rays	Rotation-invariant	Lower accuracy than CT-based	73%	NIH Chest X-ray dataset (Kaggle)
ML for Chest X-ray Interpretation (Ahmad et al.) [12]	Review of DL models	Reviews chest X- ray classifiers	Real-world insights	Mostly retrospective	Up to 94%	NIH, CheXpert, MIMIC-CXR, etc.
Abnormality Detection (Khalil, 2024) [8]	Gabor + DNN	X-ray anomaly detection	Textural features improved	Moder ate accuracy	79%	NIH Chest X-ray
Pathologists in Lung Disease (Calabrese et al.) [2]	Expert review	Describes diagnostic role	Clinical relevance	No AI/data used	N/A	N/A
Lung Diseases DL (Salih et al.) [7]	CNN, DNN, GAN	Detects pneumonia, TB, COVID-19	Covers multiple disease s	Needs large data	Not specified	NIH, COVIDx, open
Pulmonary Disease Multi-Classification (Rahul et al.) [4]	CNN, MobileNet, Inception	Multi-disease classification	High accuracy	Deep model complexity	95.6%	Combined datasets
Respiratory Disease Detection (2023) [1]	CNN	Detects pneumonia/COVI D	Fast diagnosis	Limited dataset info	~95%	Not specified
Detection Using CT & X-Ray [11]	CNN	COVID-19 and Effusion	Combines CT & X-ray	Model detail limited	Not stated	Custom CT set

V. Challenges and Issues

While Artificial Intelligence (AI) has shown impressive potential in detecting lung conditions using chest X-ray images, and while many of challenges still limit the implementation of AI for common use in the real-world clinical space, significant challenges remain which limit broader use in real world clinical environments.

A. Imbalanced Datasets and Limited Data Availability

The most significant limitation is the very data itself. The majority of currently available datasets are extremely imbalanced, either containing an enormous number of normal or commonly seen cases and a small number of rare but important conditions, such as lung cancer or TB. This imbalance can create AI models that are biased to only predict a more prevalent disorder, which can make rare but critical disorders more difficult to identify. The other factor is the lack of good, labeled medical images—especially for rare diseases—that are required to build accurate and reliable models.

B. Lack of Model Generalization

AI models often perform well in controlled study contexts but tend to lose performance when executed in real-world situations; machine performance decreases when the specific performance determining factors become different than those in the original context. The performance of AI models may decrease due to the different imaging equipment, var

C. Limited Explainability and Trust

Most deep learning models are complex systems that make decision-making opaque, and this opacity is a source of problems for medical practitioners.
While heatmaps and feature attribution methods reveal how the model made a decision, there is nothing that says they will be transparent for medical practitioners.
This can cause physicians to distrust artificial intelligence (AI) diagnoses when the reasoning for the output of the model is not transparent.

D. Real-World Implementation Challenges

Bringing AI to life in the healthcare setting has its own practical problems. Many sophisticated AI systems require high-performance computing hardware, and potentially have training assets unavailable to smaller hospitals or locations with deficient healthcare infrastructure.

E. Lack of Clinical Context in Diagnosis

AI models that utilize imaging data exclusively might be overlooking important context: symptomatology, relevant past medical history, age, or test results. As they use a narrower span of data, errors can occur in identifying patient disease, especially when developing a differential diagnosis for diseases that will look identical on an X-ray (pneumonia, COVID-19, or TB as examples). More clinical data may substantially improve AI diagnostic accuracy.

F. Inaccurate or Inconsistent Labeling

Another issue is the reliability of the labels in the training dataset. Training datasets of some medical images will use labels that are automatically generated from radiologic scans or from clinical comments. If there are labeling errors, it can mislead the model when it is trained, producing inaccurate predictions and these labels are not always accurate.

G. Ethical, Legal, and Regulatory Hurdles

The integration of AI in health care poses serious challenges regarding patient privacy, potential biases in algorithms, and some sort of dependence on laws and regulated environment. In order to ethically use AI technology, openness, equity across all patients, attention to health care regulations, etc., has to occur. Failure to address these challenges means social acceptance, or implementation of AI systems into clinical development would be limited to select, niche practices.

VI. Conclusion

The hybrid deep learning model, which the authors suggest based on the integration of CNN and LSTM architectures, presents a promising avenue for early and precise diagnosis of occupational lung conditions like pneumoconiosis and COPD among coal miners. By combining image-derived features from chest X-rays with contextual health information—such as years of service, smoking history, BMI, and duration of exposure—the system seeks to address the shortcomings of current models that tend to use imaging data exclusively.
The model is not only intended to be highly accurate, but also to be deployed practically in real-world, resource-limited environments, where radiological skill and sophisticated computing resources could be unavailable. By incorporating explainable AI methods, the system also seeks to establish trust with healthcare professionals and provide clinical transparency.
By overcoming typical issues like dataset imbalance, generalizability problems, and low interpretability, the system presented has potential to emerge as a trustworthy diagnostic support system. Its capacity to identify early disease symptoms can provide timely interventions, eventually enhancing the quality of life and health condition of the workers in dangerous workplaces like mining..

Discussion

The suggested system will create a deep learning hybrid architecture for the early detection and classification of occupational lung diseases—particularly pneumoconiosis and COPD—based on chest X-ray images and additional health data obtained from coal mine workers. Following recent advances in AI diagnostics, the system combines Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal pattern learning, particularly when there is sequential or longitudinal data such as periodic health screening [7].
The approach is motivated from high-accuracy models that have been discussed in literature, such as Xception, DenseNet121, and ensemble approaches, which have exhibited high potential for classifying disease based on CT and X-ray [10]. Yet, the majority of such models are single-modality and disease-specific, and usually, they don't generalize well to different clinical conditions [3].
Our system stands out by:

- Blending spatial and temporal features to enable more comprehensive representations of disease progression and severity.
- Combining non-image information like working conditions, symptoms reported, smoking status, blood pressure, and BMI—rendering it a multi-modal diagnostic tool.
- Applying data augmentation and optimization methods (such as horizontal flip, zoom, and rotation) to enhance model robustness and generalizability.
- Overcoming generalization issues by incorporating data from heterogeneous population groups and from various imaging sources (such as open-source datasets and hospital records).

REFERENCES

[1] Laxmibai, & Patil, V. (2023). “Enhancing respiratory disease diagnosis through deep learning: A CNN-based approach for image classification”. 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS), pp. 1–6. IEEE. https://doi.org/10.1109/ICIICS59993.2023.10421028

[2] Preethi, S., Raghul, M., Ezhillin Freeda, S., & Surya Prasath, J. P. (2024). “Enhanced approach to detect disease in the respiratory system.” 2024 International Conference on Circuit Power and Computing Technologies (ICCPCT), IEEE, pp. 1–6. https://doi.org/10.1109/ICCPCT61902.2024.10673218

[3] Chetan Reddy V., Naveen Kumar V., Padma Sai Y., Spurthi G., and Mahesh A., "Multi-Classification of Respiratory Diseases using Deep Learning," Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS 2023), IEEE, pp. 273–274, 2023, https://doi.org/10.1109/ICSCSS57650.2023.10169597

[4] R. V. Rahul, O. R., V. P. Balaji, S. V. Mudunuri, T. Singh, A. Tripathi, P. Duraisamy, and S. S. Shivakumar, “Pulmonary disease multiclassification of chest X-ray images using deep learning techniques,” in Proc. IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2024, pp. 1–8. [Online]. Available: https://doi.org/10.1109/RAICS61201.2024.10689740

[5] D. R. Rathod and S. M. Rathod, [7]” in Proc. 2023 2nd Int. Conf. on Trends in Electrical, Electronics, and Computer Engineering (TEECCON), 2023, pp. 117–121. [Online]. Available: https://doi.org/10.1109/TEECCON59234.2023.10335887

[6] L. S. Safwat and A. M. Nagy, “Prediction of Silicosis in Iron and Steel Workers Using Artificial Intelligence,” Journal of Southwest Jiaotong University, vol. 58, no. 2, pp. 167–175, Apr. 2023. [Online]. Available: https://doi.org/10.35741/issn.0258-2724.58.2.15

[7] F. A. Mostafa, L. A. Elrefaei, M. M. Fouda, and A. Hossam, "Diagnosis of Lung Diseases from Chest X-Ray Images Using Different Fusion Techniques," 2023 11th International Conference on Information and Communication Technology (ICoICT), IEEE, pp. 429–433, 2023. http://doi.org/10.1109/ICoICT58202.2023.10262761

[8] M. S. Khalil, "Deducting Abnormalities in Chest X-Rays using Gabor Filters and Deep Neural Network (DNN),"Jurnal Kejuruteraan, vol. 36, no. 5, pp. 1965–1972, 2024.http://doi.org/10.17576/jkukm-2024-36(5)-16

[9] E. K. Austin, C. James, and J. Tessier, “Early detection methods for silicosis in Australia and internationally: A review of the literature,” Int. J. Environ. Res. Public Health, vol. 18, no. 15, p. 8123, Jul. 2021. [Online]. Available: https://doi.org/10.3390/ijerph18158123

[10] A. Nainar M., P. Kumar A., S. Ruban D., R. Raj K., and R. S., "Detection of Pulmonary Diseases in Lung Using CT & X-Ray Images by CNN," International Journal for Multidisciplinary Research (IJFMR), vol. 6, no. 2, pp. 1–4, Mar.–Apr. 2024. http://www.ijfmr.com

[11] H. K. Ahmad et al., “Machine learning augmented interpretation of chest X-rays: A systematic review,” Diagnostics, vol. 13, no. 4, p. 743, Feb. 2023. [Online]. Available: https://doi.org/10.3390/diagnostics13040743

[12] 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, Article no. 100391, Jul. 2020. [Online]. Available: https://doi.org/10.1016/j.imu.2020.100391