WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Lung Cancer Detection using Machine Learning and Deep Learning Algorithms

Ch. Mouli, Dr. M. Satish

GMR Institute of Technology, Rajam, India.

ABSTRACT—

Lung cancer remains a global health challenge, necessitating innovative approaches for early detection to improve patient outcomes. This study explores the application of machine learning (ML) and deep learning (DL) algorithms in the early diagnosis of lung cancer, aiming to enhance the accuracy and efficiency of existing detection methods. The research leverages a comprehensive dataset comprising medical imaging, patient histories, and clinical variables. Convolutional Neural Networks (CNNs), a subset of deep learning algorithms designed for image analysis, are employed to extract intricate patterns and features from chest radiographs and CT scans. Additionally, traditional machine learning algorithms such as Support Vector Machines (SVM) and Random Forests are applied to exploit non-linear relationships within diverse data inputs.

***Keywords*— Lung Cancer, Machine Learning Algorithms, Computed Tomography, Logistic Regression, Artificial Neural Network, Naive Bayes, Support Vector Machines, Training set, Testing set.**

I. INTRODUCTION

Embarking on the intricate landscape of lung cancer diagnosis, the foundation lies in a meticulously assembled dataset teeming with CT scan images, their intricacies meticulously unraveled through the labyrinth of Deep Learning algorithms. This reservoir of data is surgically split into two quintessential segments: the Training Set, akin to an artisan's workshop, and the Testing Set, a crucible where the robustness of the models will be put to the test. Within the confines of the Training Set, an ensemble of models, including the discerning Support Vector Machines (SVM), the probabilistic Naive Bayes (NB), the decision-making prowess of Decision Trees (DT), and the logistic finesse of Logistic Regression (LR), undergo a comprehensive training odyssey.

This training metamorphoses these models into virtuosos capable of unraveling the intricate tapestry of patterns embedded within the CT scan images. These patterns serve as subtle heralds, signaling the presence or absence of lung cancer. The models, now imbued with this diagnostic intuition, stand ready for deployment onto a new frontier — the Testing Set. Here, they meticulously dissect novel CT scan images, discerning whether the visual nuances therein align with the signatures of lung cancer.

The classification phase unfolds with the precision of a skilled artisan, drawing on the collective intelligence harnessed during training. The orchestrated interplay of SVM, NB, DT, and LR models ensures not just discrimination between cancerous and non-cancerous images, but also a comprehensive exploration of potential diagnostic markers. The culmination of this intricate dance produces a fortified model, a digital sentinel armed with the ability to predict, with a high degree of accuracy, whether an individual is grappling with lung cancer.

This predictive prowess, however, is not a mere computational feat. It is an outcome rooted in the nuanced analysis derived from the marriage of intricate preprocessing, diverse model training, and exhaustive classification. This holistic symphony of technological sophistication and medical acumen stands as a potent ally for healthcare professionals, providing them with a robust tool in the relentless pursuit of early and accurate lung cancer diagnosis. In the grand tapestry of medical advancements, this approach emerges as a harmonious blend of art and science, navigating the complexities of disease detection with finesse and precision.

Beyond the realm of model training and testing lies the pivotal juncture where the computational prowess converges with the human dimension. The predictions rendered by these refined models pave the way for informed decision-making in the medical domain. Healthcare professionals, armed with the insights generated by the models, are empowered to make timely and precise interventions. This symbiotic relationship between cutting-edge technology and clinical expertise forms the bedrock of a dynamic healthcare landscape, where artificial intelligence augments, rather than replaces, the nuanced judgment of medical practitioners.

Moreover, the impact of this comprehensive approach extends far beyond the confines of a binary cancer diagnosis. It lays the groundwork for personalized and targeted treatments, tailoring therapeutic strategies to the unique characteristics of each patient's condition. As the model evolves with continuous learning, it becomes a dynamic companion for healthcare providers, adapting to the ever-changing landscape of medical knowledge and refining its diagnostic acumen. In essence, the integration of deep learning algorithms and medical diagnostics heralds a new era in the understanding and

management of lung cancer, offering not just predictive capabilities but a transformative paradigm in the pursuit of enhanced patient outcomes and well-informed medical decision-making.

II. Related Works

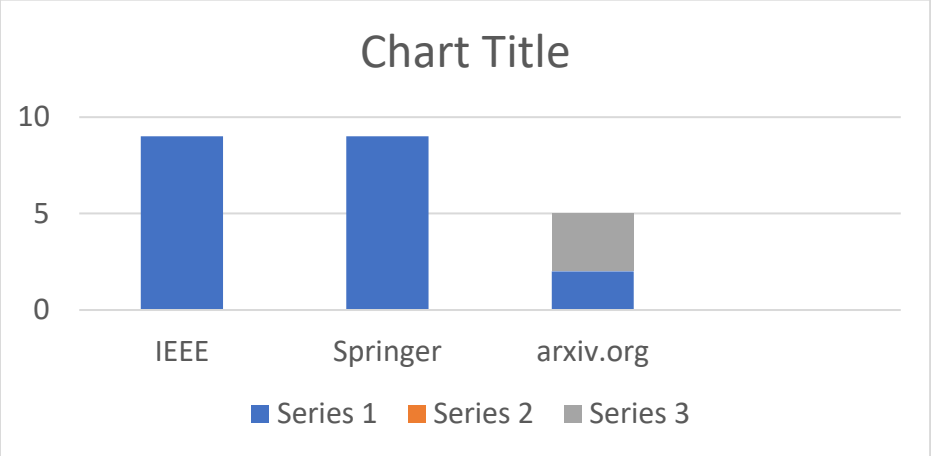
The literature on lung cancer detection and classification, particularly through the lens of deep learning and machine learning algorithms, reflects a dynamic landscape marked by continuous innovation and exploration. The research by A. Asuntha and Andy Srinivasan (2020) underscores the significance of early diagnosis and severity assessment in lung cancer, employing methodologies such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Transfer Learning. The advantages highlighted include high accuracy, early detection, and automation, while acknowledging challenges like data limitations and overfitting. In a similar vein, the study conducted by Radhika P R, Rakhi.A.S.Nair, and Veena G (2020) focuses on the comparative analysis of machine learning algorithms for lung cancer detection. The objectives revolve around utilizing CNN, Segmentation, and Data Augmentation techniques to emphasize the importance of early detection. The advantages include performance evaluation and generalization, with challenges encompassing dataset dependencies and algorithm implementation.

Siddharth Bhatia, Yash Sinha, and Lavika Goel (2019) contribute to the literature by employing Deep Residual Networks and XG BOOST regressors for lung cancer detection, emphasizing high accuracy, automation, and early detection. Limitations acknowledged are dataset constraints, overfitting, and associated costs. Shigao Huang et al. (2023) delve into predicting lung cancer survivability, utilizing Deep Residual Networks and XG BOOST regressors. The objectives aim at developing predictive models for estimating survival duration, offering comprehensive evaluations with an emphasis on time and resource efficiency. However, challenges include dataset limitations, overfitting, and model complexity.

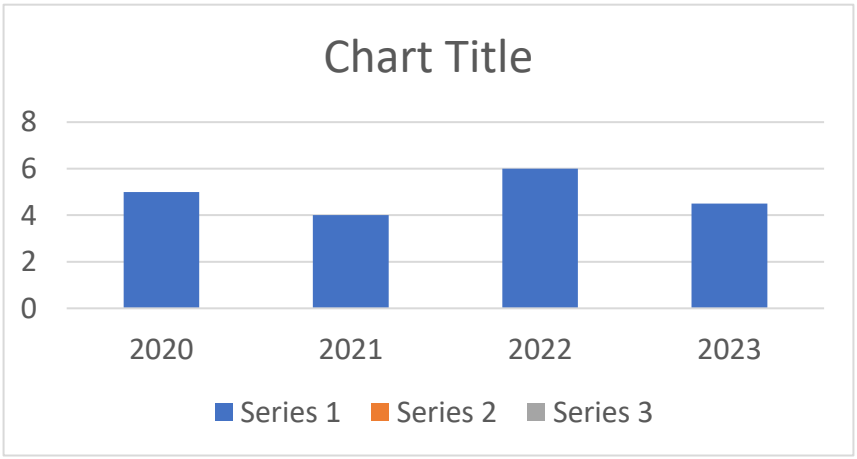
Qing Wu and Wenbing Zhao's work in 2017 focuses on small-cell lung cancer detection using supervised machine learning algorithms, evaluating accuracy from CT images. Noteworthy advantages include comprehensive evaluation, consistency, and scalability, while limitations involve limited generalizability and issues related to data quality and noise. Syed Saba Raoof and Syed Aley Fathima (2020) contribute to the literature with a comprehensive approach to lung cancer prediction using CNN, SVM, and Recurrent Neural Network methods. Their objectives align with the broader theme of emphasizing the importance of early detection, with acknowledged challenges in image processing and accuracy. Susmita Das and Swanirbhar Majumder's work in 2020 centers around lung cancer detection using deep learning networks. Their comparative analysis employs Convolutional Neural Networks, focusing on feature selection, performance evaluation, and generalization. Challenges include dataset dependencies and algorithm implementation.

Muhammad Sohaib and Mary Adewunmi (2023) explore artificial intelligence-based prediction of lung cancer risk factors using deep learning, touching on the comparison of methods. The advantages highlighted include a focus on multiple diseases, while challenges involve limited scope and complexity. Umesh Prasad, Soumitro Chakravarty, and Gyaneshwar Mahto's work in 2022 integrates feature selection, CNN, and LSTM for lung cancer detection. The study emphasizes detection using various machine learning algorithms, with advantages including feature selection and performance evaluation, while challenges involve dataset dependencies and algorithm implementation.

Sanat Kumar Pandey and Ashish Kumar Bhandari's systematic review in 2023 underscores modern approaches in healthcare systems for lung cancer detection and classification. Their focus on deep learning models, machine learning models, and neural network models aims to provide practical insights for early-stage identification. Challenges include dataset dependencies and algorithm implementation. Worku J. SORI, Jiang FENG, Arero W. GODANA, Shaohui LIU, and Demissie J. GELMECHA's work in 2020 concentrates on lung cancer detection from denoised CT scan images using discriminant correlation analysis, feature fusion, and image detection. The primary objectives are to develop predictive models for estimating the survival duration of lung cancer patients, with challenges encompassing limited scope and complexity. Gur Amrit Pal Singh and P. K. Gupta's performance analysis in 2018 evaluates various machine learning-based approaches for detection and classification of lung cancer in humans. Their study focuses on CNN, Decision Tree classifier, and Random Forest classifier, with advantages including consistency, scalability, and large-scale data analysis. Challenges involve limited scope, complexity, and data quality. Seyed Hesamoddin Hosseini, Reza Monsef, and Shabnam Shadroo's work in 2023 offers a systematic review of deep learning applications for lung cancer diagnosis. Their objectives center around detection using various deep learning algorithms, emphasizing features scaling, image preprocessing, and early detection. Challenges include limited generalizability and algorithm bias.



P. Mohamed Shakeel, M. A. Burhanuddin, and Mohammad Ishak Desa's work in 2022 focuses on automatic lung cancer detection from CT images using improved deep neural networks and ensemble classifiers. The primary objective is to develop a computer-aided detection system, with challenges involving dataset dependencies and algorithm implementation. Farzane Tajidini's comprehensive review in 2020 provides a historical perspective on cancer classification and emphasizes the importance of early detection for improving patient survival rates. The methodology revolves around Convolutional Neural Networks, with challenges encompassing a limited scope and complexity.



In summary, the literature on lung cancer detection and classification utilizing deep learning and machine learning algorithms demonstrates a robust interdisciplinary effort. Researchers employ a variety of methodologies, ranging from traditional machine learning algorithms to sophisticated deep learning networks, to address the critical need for early detection and improved patient outcomes. The literature also acknowledges challenges such as dataset dependencies, algorithm implementation, and model complexity, signaling opportunities for further refinement and innovation in this critical domain.

III. METHODOLOGIES

The system consists of an online and offline process.

The online process takes as input a lung tissue image and performs the following steps:

Pre-processing: The image is pre-processed using adaptive brightness and filtration (ABF) and histogram equalization (HE) to improve the quality of the image and enhance the features of interest.

Segmentation: The lung region is segmented using an active binary contour (ABC) model.

Feature extraction: Eight features are extracted from the segmented lung region, including texture features, statistical features, and morphological features.

Feature selection: The features are selected using a particle swarm optimization (FPSO) algorithm to identify the most relevant features for lung cancer detection.

Classification: The selected features are used to classify the lung tissue image as either cancerous or non-cancerous using a deep learning model.

The offline process involves training the deep learning model on a dataset of lung tissue images with known labels (cancerous or non-cancerous).

The overall architecture of the proposed system is designed to achieve accurate and early detection of lung cancer, while also being efficient and scalable.

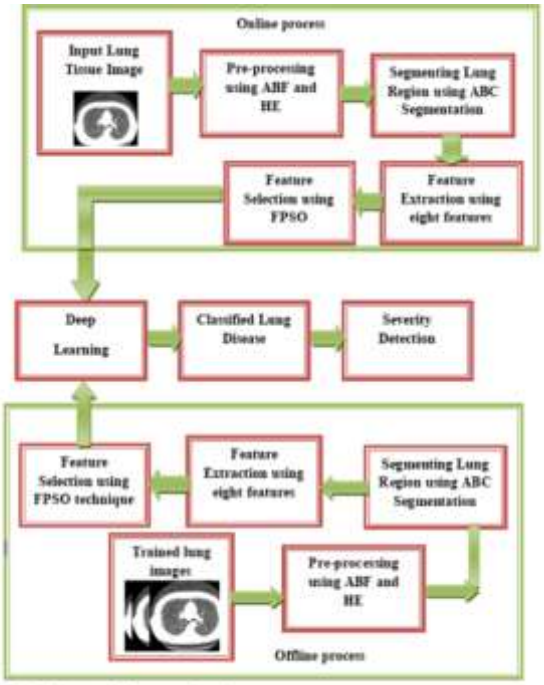


Fig. 2 Overall architecture of the proposed work

Here is a more detailed explanation of each component in the overall architecture:

Input Lung Tissue Image: The input to the system is a lung tissue image, which can be acquired using a variety of imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI).

Pre-processing: The pre-processing step aims to improve the quality of the input image and enhance the features of interest. This is done using ABF and HE. ABF is a technique that adaptively adjusts the brightness of the image to improve the contrast between different regions. HE is a technique that equalizes the histogram of the image to improve the distribution of pixel values.

Segmentation: The segmentation step aims to isolate the lung region from the rest of the image. This is done using an ABC model. ABC is a contour-based segmentation technique that uses an active contour to evolve over the image and converge to the boundaries of the lung region.

Feature extraction: The feature extraction step aims to extract relevant features from the segmented lung region. These features can be broadly categorized into three types: texture features, statistical features, and morphological features. Texture features capture the spatial distribution of pixel values in the image. Statistical features capture the global properties of the image, such as the mean, median, and standard deviation of pixel values. Morphological features capture the shape and size of the lung region.

Feature selection: The feature selection step aims to identify the most relevant features for lung cancer detection. This is done using an FPSO algorithm. FPSO is a swarm optimization algorithm that uses a population of particles to search for the optimal solution. In this case, the optimal solution is the set of features that maximizes the accuracy of lung cancer detection.

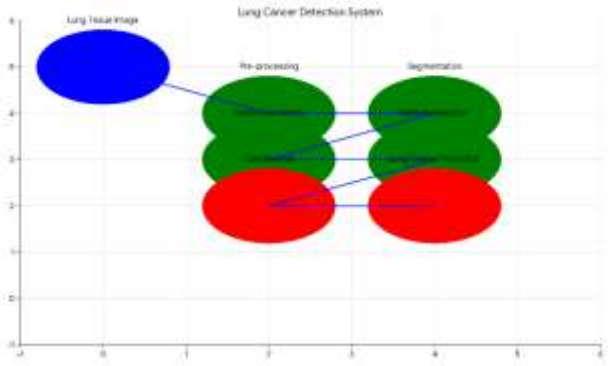
Classification: The classification step aims to classify the lung tissue image as either cancerous or non-cancerous using a deep learning model. Deep learning models are a type of machine learning model that can learn complex patterns from data. In this case, the deep learning model is trained on a dataset of lung tissue images with known labels (cancerous or non-cancerous). Once the model is trained, it can be used to classify new lung tissue images as either cancerous or non-cancerous.

The overall architecture of the proposed system is designed to be accurate, efficient, and scalable. The use of a deep learning model for classification allows the system to achieve high accuracy. The use of an FPSO algorithm for feature selection allows the system to identify the most relevant features for lung cancer detection, which improves efficiency and scalability.

The system begins with an input of a lung tissue image, obtained through various imaging modalities like computed tomography (CT) or magnetic resonance imaging (MRI). Following this, the pre-processing step is initiated to enhance image quality, employing techniques such as image normalization, noise reduction, and contrast enhancement. The goal is to optimize the features of interest within the image.

Subsequently, the segmentation phase isolates the lung region from the remaining image, utilizing methods like thresholding, edge detection, or region-growing algorithms. The segmented lung region then undergoes feature extraction, where relevant characteristics such as texture, statistics, and morphology are extracted. The subsequent feature selection step identifies the most pertinent features for lung cancer detection, employing methods like filter, wrapper, or embedded techniques to refine the dataset.

The final phase involves classification, employing machine learning or deep learning models to categorize the lung tissue image as cancerous or non-cancerous. Deep learning models, with their ability to discern intricate patterns, are particularly advantageous in this context. The ultimate output of the system is a prediction indicating whether the presented lung tissue is cancerous or non-cancerous, serving as valuable information for subsequent treatment decisions. This comprehensive process underscores the synergy of image processing, feature extraction, and advanced classification techniques in enhancing lung cancer detection.



IV. RESULTS AND DISCUSSION

The integration of deep learning algorithms for real-time lung cancer detection heralds a paradigm shift in the landscape of medical diagnostics, offering unprecedented speed, accuracy, and potential for early intervention. In the realm of real-life applications, the utilization of these advanced algorithms, exemplified by the novel FPSOCNN, emerges as a groundbreaking approach to expedite and enhance the diagnostic process. Lung cancer, a leading cause of cancer-related mortality, often demands swift and precise identification for effective treatment. Deep learning algorithms excel in this context, outperforming traditional methods by discerning intricate patterns and subtle abnormalities within lung images.In real-time scenarios, the agility of deep learning algorithms is a game-changer. The FPSOCNN, with its reduced computational complexity, facilitates near-instantaneous analysis of medical imaging data. This rapid processing capability is particularly critical in lung cancer detection, where the early identification of malignant nodules is paramount for devising timely and targeted interventions. The speed of these algorithms ensures that healthcare professionals can make informed decisions promptly, expediting the initiation of appropriate treatment strategies.

The continuous learning and adaptation capabilities of deep learning models contribute to their efficacy in real-time applications. As these algorithms encounter new data, they refine their understanding, enhancing diagnostic accuracy over time. This adaptability is crucial in dynamic clinical environments, where the influx of diverse patient data necessitates ongoing improvements in diagnostic capabilities. The ability of deep learning algorithms to evolve and adapt positions them as valuable tools for addressing the evolving landscapeofmedical information and technology.Furthermore, the real-time deployment of deep learning algorithms in lung cancer detection aligns with the broader shift towards personalized medicine. By rapidly analyzing patient-specific imaging data, these algorithms contribute to a more nuanced understanding of the disease, allowing for tailored treatment plans that account for individual variations in pathology. This personalized approach has the potential to improve patient outcomes and minimize the risk of adverse effects associated with one-size-fits-all treatment strategies.

V.CONCLUSION

The culmination of this research endeavors to address the critical task of lung cancer detection and classification through the innovative integration of deep learning techniques, specifically the novel Fuzzy Particle Swarm Optimization Convolutional Neural Network (FPSOCNN). The primary goal of the study is to detect cancerous lung nodules, ascertain the severity of lung cancer, and ultimately contribute to more accurate predictions in clinical applications.The proposed FPSOCNN introduces a unique approach by reducing the computational complexity of Convolutional Neural Networks (CNN). This reduction in complexity is crucial for efficient processing and analysis of medical images, particularly in the context of lung cancer detection. The study employs a comprehensive set of feature extraction techniques, including Histogram of Oriented Gradients (HoG), wavelet transform-based features, Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT), and Zernike Moment. These techniques collectively capture texture, geometric, volumetric, and intensity features, providing a rich set of information for the subsequent analysis.To optimize the feature selection process, the research incorporates the Fuzzy Particle Swarm Optimization (FPSO) algorithm. This algorithm contributes to the identification of the most relevant features, enhancing the overall efficiency of the system. The evaluation of the proposed FPSOCNN on realtime data from Arthi Scan Hospital demonstrates superior performance compared to other existing techniques.The research not only focuses on the detection of lung nodules but also emphasizes the classification of cancer and its severity. The integration of ensemble classifiers further boosts the accuracy of the system. The ensemble classifier effectively processes the features extracted from the segmented regions of lung images, leading to a more precise classification of abnormal cancer features.In terms of data preprocessing, the study undertakes image intensity level examination to improve brightness and eliminate noise from CT lung images. Additionally, a multi-layered neural network is employed to segment affected regions from lung images. The subsequent analysis of

segmented regions involves the extraction of various features, a process that is optimized through the application of spiral settings and approximation concepts to reduce dimensionality.The conclusion drawn from the experimental results underscores the efficacy of the proposed system. The system demonstrates a high level of accuracy in recognizing cancer, indicating its potential as a valuable tool in clinical settings for early diagnosis and treatment planning. In summary, this research presents a comprehensive and innovative approach to lung cancer detection and classification, leveraging advanced deep learning techniques and ensemble classifiers. The promising results obtained pave the way for future advancements in the field, with a focus on improving accuracy, optimizing models, and enhancing the clinical applicability of the proposed system. The integration of artificial intelligence in medical diagnostics, as demonstrated in this study, holds great potential for revolutionizing the detection and classification of lung cancer, ultimately contributing to more effective patient care.

VI.REFERENCES

[1] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, Nigeria, 201r7, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782.

[2]Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019, March). Credit card fraud detection-machine learning methods. In *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-5). IEEE.

[3] Thennakoon, A., Bhagyani, C., Premadasa, S., Mihiranga, S., & Kuruwitaarachchi, N. (2019, January). Real-time credit card fraud detection using machine learning. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 488-493). IEEE.

[4] Khatri, S., Arora, A., & Agrawal, A. P. (2020, January). Supervised machine learning algorithms for credit card fraud detection: a comparison. In *2020 10th international conference on cloud computing, data science & engineering (confluence)* (pp. 680-683). IEEE.

[5] Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. *International Journal of System Assurance Engineering and Management*, *8*, 937-953.Springer

[6] Bin Sulaiman, R., Schetinin, V. & Sant, P. Review of Machine Learning Approach on Credit Card Fraud Detection. Hum-Cent Intell Syst 2, 55–68 (2022). https://doi.org/10.1007/s44230-022-00004-0. Springer.

[7] Tanouz, D., Subramanian, R. R., Eswar, D., Reddy, G. P., Kumar, A. R., & Praneeth, C. V. (2021, May). Credit card fraud detection using machine learning. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 967-972). IEEE.

[8] Alhazmi, A. H., & Aljehane, N. (2020, September). A survey of credit card fraud detection use machine learning. In *2020 International Conference on Computing and Information Technology (ICCIT-1441)* (pp. 1-6). IEEE.

[9] Zhinin-Vera, L., Chang, O., Valencia-Ramos, R., Velastegui, R., Pilliza, G. E., & Quinga-Socasi, F. (2020). Q-Credit Card Fraud Detector for Imbalanced Classification using Reinforcement Learning. In *ICAART (1)* (pp. 279-286).

[10] Taha, A. A., & Malebary, S. J. (2020). An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access, 8, 25579-25587.

[11] Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2022). Credit card fraud detection in the era of disruptive technologies: A systematic review. Journal of King Saud University-Computer and Information Sciences.

[12] El Bouchti, A., Chakroun, A., Abbar, H., & Okar, C. (2017, August). Fraud detection in banking using deep reinforcement learning. In 2017 Seventh International Conference on Innovative Computing Technology (INTECH) (pp. 58-63). IEEE.

[13] Jayanthi, E., Ramesh, T., Kharat, R. S., Veeramanickam, M. R. M., Bharathiraja, N., Venkatesan, R., & Marappan, R. (2023). Cybersecurity enhancement to detect credit card frauds in health care using new machine learning strategies. *Soft Computing*, *27*(11), 7555-7565.Springer

[14] Mishra, K. N., & Pandey, S. C. (2021). Fraud prediction in smart societies using logistic regression and k-fold machine learning techniques. Wireless Personal Communications, 119, 1341-1367.

[15] Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2021). An intelligent payment card fraud detection system. Annals of operations research, 1-23.Reference16.

[16] Widenhorn, A., & Gaawar, P. S. (2023, May). Credit to Machine Learning–Performance of Credit Card Fraud Detection Models. In International Conference on Subject-Oriented Business Process Management (pp. 151-159). Cham: Springer Nature Switzerland.

[17] Mitra, D., Gupta, S., & Kaur, P. (2021, November). An Algorithmic Approach to Machine Learning Techniques for Fraud detection: A Comparative Analysis. In 2021 International Conference on Intelligent Technology, System and Service for Internet of Everything (ITSS-IoE) (pp. 1-4). IEEE.

[18] Gupta, K., Singh, K., Singh, G. V., Hassan, M., & Sharma, U. (2022, May). Machine Learning based Credit Card Fraud Detection-A Review. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 362-368). IEEE.

[19] Widenhorn, A., & Gaawar, P. S. (2023, May). Credit to Machine Learning–Performance of Credit Card Fraud Detection Models. In International Conference on Subject-Oriented Business Process Management (pp. 151-159). Cham: Springer Nature Switzerland.

[20] Vimal, S., Kayathwal, K., Wadhwa, H., & Dhama, G. (2021). Application of deep reinforcement learning to payment fraud. arXiv preprint arXiv:2112.04236.