Grenze International Journal of Engineering and Technology, June Issue

Enhancing Healthcare through the Symbiosis of Blockchain and Machine Learning in Preprocessing Skin Disease Images: A Review

Ms. Sonali Rokade[1] and Dr Nilamadhab Mishra[2]
[1-2]School of Computer Science and Engineering, Vellore Institute of Technology, Bhopal, India.
Email: rokade.shashikant2019@vitbhopal.ac.in, nilamadhab.mishra@vitbhopal.ac.in

***Abstract*— Recent advancements in skin disease identification leverage machine learning for automated diagnosis. However, safeguarding the integrity of sensitive medical data remains a top priority. This paper explores the integration of machine learning and blockchain in skin disease identification preprocessing, promising trust, transparency, and accuracy in dermatological diagnosis. The preprocessing workflow starts with gathering a diverse, labeled skin image dataset, followed by data cleaning and augmentation to enhance quality and quantity. Standardization methods like resizing and normalization ensure data consistency, while addressing class imbalances prevents model bias. Blockchain technology is introduced to secure data integrity. Each skin image and its metadata are encrypted and timestamped on a decentralized blockchain, guaranteeing tamper-proof authenticity. Machine learning models, typically based on CNNs, extract features and classify diseases, benefiting from high-quality preprocessed data. During training, blockchain securely records model parameters and progress, ensuring transparency and auditability, crucial for medical accountability and compliance.This fusion of machine learning and blockchain enhances diagnostic accuracy while addressing data privacy and security concerns, potentially revolutionizing dermatological healthcare with a trustworthy, automated diagnosis system.**

***Index Terms*— Skin disease identification, Machine learning, Blockchain technology, Data integrity, Dermatological diagnosis**

I. INTRODUCTION

Skin diseases present a substantial public health challenge, impacting millions of individuals globally. Ensuring timely and precise diagnosis is crucial for effective treatment and care. Nonetheless, the domain of dermatological healthcare has grappled with persistent challenges concerning data integrity, privacy, and accountability, hindering advancements in automated diagnosis. The recent progress in technology, specifically the integration of machine learning and blockchain, offers the potential to tackle these challenges and introduce a fresh era marked by trust, transparency, and precision in the identification of skin diseases.
This study investigates the groundbreaking fusion of machine learning and blockchain technologies in the preprocessing phase of skin disease identification. The preprocessing pipeline begins with the collection of a diverse and labeled skin image dataset, followed by data cleaning and augmentation to enhance data quality and

Grenze ID: 01.GIJET.10.2.102

quantity. Standardization techniques, such as resizing and normalization, ensure uniformity in data representation, while addressing class imbalances prevents model bias. Blockchain technology is introduced as a foundational element to guarantee data integrity and security throughout the preprocessing pipeline. Each skin image, along with its associated metadata, is encrypted and timestamped on a decentralized blockchain network, creating an immutable ledger that ensures data remains tamper-proof and authentic, fostering trust among stakeholders. Blockchain, as a decentralized peer-to-peer network, eliminates the need for intermediaries like payment apps such as PayPal and Alipay. This, in turn, enhances operational efficiency and lowers transaction expenses. [1].

Machine learning models, often built on convolutional neural networks (CNNs), play a central role in feature extraction and classification tasks. The availability of high- quality, preprocessed data enhances the accuracy and reliability of these models in diagnosing a wide range of skin diseases.

Furthermore, the role of blockchain technology extends into the training phase, securely recording model parameters and training progress. This approach ensures that model updates are transparent, auditable, and traceable—a crucial aspect for medical accountability and regulatory compliance.

The fusion of machine learning and blockchain technology in skin disease identification preprocessing not only promises to enhance the accuracy of diagnosis but also addresses critical concerns regarding data privacy, security, and accountability. This approach holds the potential to revolutionize dermatological healthcare, offering a trustworthy and efficient means of automated skin disease diagnosis, ultimately improving patient outcomes and healthcare efficiency. This research sets out to explore, analyze, and substantiate these promising developments [1].

A broad spectrum of conditions falls under the category of skin diseases, ranging from common issues such as acne to more severe disorders like melanoma. Timely and precise diagnosis is often hampered by the shortage of dermatologists, leading to delayed treatment and increased healthcare costs.

the fusion of machine learning and blockchain is positioned as a revolutionary approach to dermatological healthcare. The proposed system not only enhances diagnostic accuracy but also aims to address data privacy and security concerns, offering a trustworthy and automated diagnosis system for skin diseases. The abstract suggests a potential paradigm shift in dermatological healthcare through the integration of these advanced technologies.

II. RELATED WORK

The impact of weather on symptoms in individuals with chronic diseases has been acknowledged for over 2000 years, tracing back to the time of Hippocrates. This study highlights that heightened relative humidity, increased wind speed, and lower atmospheric pressure correlate with increased pain severity in individuals with long-term pain conditions, with relative humidity playing the most substantial role. Notably, the influence of weather on pain is not entirely explained by its day-to-day effects on mood or physical activity. The combination of specific weather variables is recognized as potentially raising the odds of a pain event by slightly over 20% compared to an average day. Furthermore, the odds of a pain event increase by 12% per one standard deviation rise in relative humidity (9 percentage points), in contrast to a 4% decrease for pressure and a 4% increase for wind speed (11 mbar and 2 m s^{-1}, respectively). Such heightened risk may carry significance for individuals dealing with chronic pain. [1]

The global pandemic triggered by the COVID-19 virus, a member of the coronavirus ribonucleic acid virus family, has presented challenges for healthcare systems and societies worldwide. In response to this, a novel strategy that integrates blockchain and machine learning (Blockchain-ML) has been suggested for the efficient management of COVID-19 vaccine distribution. This inventive system comprises two modules: blockchain ensures the reliability of data records for vaccine traceability in the supply chain, while machine learning is applied for demand forecasting. The Blockchain-ML model strives for transparent tracing of COVID-19 vaccine distribution, optimizing allocation based on geographic locations and storage capacity. The research broadens its scope by employing machine learning techniques such as ML Regression and ML Classification to analyze datasets derived from blockchain. This comprehensive approach not only streamlines vaccine distribution but also improves the ability to forecast the most suitable COVID-19 vaccine based on available data, thereby enhancing the effectiveness of pandemic response efforts. [2]

Creating robust Machine Learning (ML) models necessitates a substantial volume of diverse training data to achieve accurate results of practical significance. Experimental findings indicate that, among Differential Privacy (DP)-based techniques in Federated Learning (FL), gradient pruning proves more effective in preserving user data privacy while minimizing the impact on model performance, although it has not been extensively explored in recent studies. The BEAS framework's evaluation, employing three widely recognized datasets and standard model configurations (such as MNIST), demonstrates its scalability with a larger participant pool (20 to 50) and attains a commendable accuracy of 90.11% on MNIST. The implementation of gradient pruning further enhances privacy protection without compromising model utility loss, addressing concerns present when employing traditional DP techniques. Notably, gradient pruning aids in reducing the susceptibility of the model to poisoning attacks, as highlighted [3]

III. METHODOLOGY

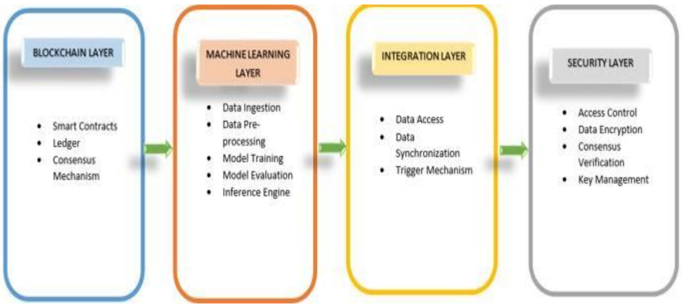

Figure 1. Blockchain-ML Integration Architecture Diagram

Figure 1. block diagram provides an overview of the components and their relationships in the architecture that combines block chain and machine learning. The Blockchain Layer interacts with the Machine Learning Layer, and the Integration Layer connects the two, allowing data to flow between them. The Security Layer ensures the integrity and security of the entire system. This is diagramming software to create a visual representation based on this textual layout [7]. Blockchain Layer:
Smart Contracts: Functioning as self-executing agreements, smart contracts encode the terms of the arrangement between the buyer and seller directly into code. These contracts enable automation and transparency in a range of transactions.
Ledger: The blockchain ledger records all transactions, providing an immutable and transparent history of actions.
Consensus Mechanism: The consensus mechanism (e.g., Proof of Work or Proof of Stake) ensures agreement on the state of the ledger.
Machine Learning Layer:
Data Ingestion: ML algorithms require data, so this layer collects and ingests data from various sources.
Data Preprocessing: Data preprocessing includes cleaning, transformation, and feature engineering to prepare the data for ML.
Model Training: This component involves training ML models using the prepared data.
Model Evaluation: After training, the models are evaluated using performance metrics.
Inference Engine: Once trained and validated, the models are deployed for inference.
Integration Layer:
Data Access: This component facilitates the secure and efficient access of data from the blockchain and other data sources for machine learning.
Data Synchronization: Ensure that data on the blockchain is synchronized with the data needed for machine learning.
Trigger Mechanism: Define how and when machine learning processes should be triggered based on blockchain events.
Security Layer:
Access Control: Implement access control mechanisms to protect sensitive data and models.
Data Encryption: Ensure data encryption in transit and at rest. Consensus Verification: Verify the integrity of data by utilizing blockchain's consensus mechanism.
Key Management: Securely manage cryptographic keys for both blockchain and machine learning components. represent this in block diagram
The figure 2, titled "Convergence of Blockchain and Machine Learning for Secure Data Analysis," illustrates how blockchain technology and machine learning intersect to enable secure data analysis. It shows how blockchain provides secure data sharing and validation, while machine learning components encompass data analysis, model training, and inference. This convergence is applicable in various fields, including healthcare, finance, supply

chain, IoT, and education, and is characterized by key features such as immutability, transparency, and decentralization, ensuring data integrity and privacy.

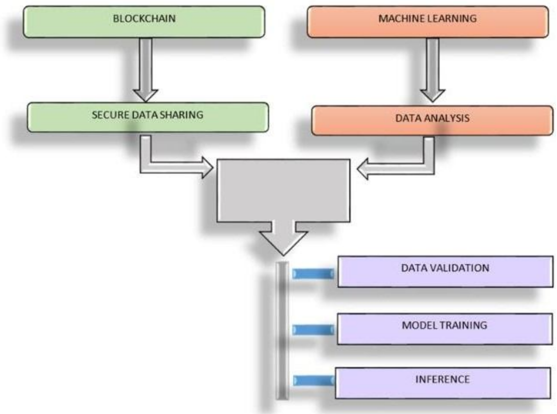

Figure 2. Synergizing Blockchain and Machine Learning: A Unified Framework for Secure Data Analytics

Features: Immutability, Transparency, Decentralization The FIGURE 2 helps convey the following:

1. *Conceptual Understanding*: It visually explains how blockchain and machine learning components work together to achieve secure data analysis. It clarifies the flow of data and processes involved in this convergence.
2. *Key Components*: It highlights the key components and their connections, such as secure data sharing, data validation, data analysis, model training, and inference.
3. *Applications:* It indicates the potential applications where this convergence can be utilized, including healthcare, finance, supply chain, IoT, and education.
4. *Key Features*: It emphasizes essential features of blockchain technology, like immutability, transparency, and decentralization, which are crucial for maintaining data integrity and privacy.

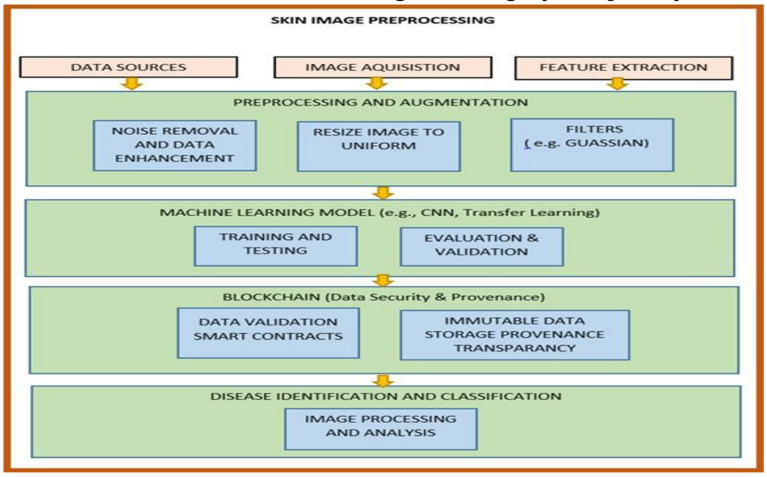

Figure 3. Integrated system for skin image preprocessing, disease identification with machine learning, and blockchain security.

Figure 3 titled "Integrated System for Skin Image Preprocessing, Disease Identification with Machine Learning, and Blockchain Security."

1. *Data Sources*: These are the various channels from which skin image data is collected. They can include medical imaging devices, mobile applications, or image repositories.
2. *Image Acquisition:* This step involves capturing skin images through devices such as cameras or medical imaging equipment. The captured images are then stored in a digital format.
3. *Feature Extraction*: In this phase, relevant features are extracted from the skin images. These features may include color, texture, and shape characteristics, which are important for identifying skin diseases.

4. *Preprocessing and Augmentation*: This component focuses on preparing the data for analysis. It includes tasks like noise removal, resizing images to a consistent format, and applying filters like Gaussian filters for enhancing image quality.
5. *Machine Learning Model*: This represents the core of the disease identification process. Machine learning models, such as Convolutional Neural Networks (CNNs), are employed for training and testing. CNNs can automatically learn and recognize patterns in skin images.
6. *Training & Testing*: Within the machine learning model, data is divided into training and testing sets. The model learns from the training data to make accurate predictions on unseen skin images from the testing set.
7. *Blockchain*: Blockchain technology is integrated into the system to ensure data security and integrity. It provides features like immutable data storage, data provenance, and transparency. Smart contracts may be employed to automate certain processes.
8. *Data Validation*: Blockchain includes mechanisms for validating the data, ensuring that it meets predefined criteria. This is essential for maintaining data integrity.
9. *Immutable Data Storage*: Blockchain stores the skin image data in a tamper-proof manner, making it resistant to unauthorized modifications.
10. *Provenance*: This refers to the tracking of data's origin and history. In the context of skin images, it ensures that the source and any changes made to the data are recorded and transparent.
11. *Disease Identification & Classification*: This is the final stage where the machine learning model, having been trained on preprocessed data, identifies and classifies skin diseases. It recognizes patterns in the images and provides diagnostic outputs.

Figure 3 illustrates the entire process from data acquisition to the secure and accurate identification of skin diseases. By integrating machine learning and blockchain, it ensures data privacy, integrity, and transparency in healthcare applications.

IV. RESULT

TABLE I. BASIC DATASET STATISTICS

Statistic	Value
count	9958
mean	51.863828
std	16.968614
min	0
25%	40
50%	50
75%	65
max	85

Table I The provided dataset statistics ON HAM10000 DATASET. These are descriptive statistics for a variable named "age" in a dataset. Descriptive statistics are used to summarize and describe the main characteristics of a dataset. Let's explain each of the statistics:

Count: The count indicates the number of non-missing or valid data points in the "age" variable. In this dataset, there are 9,958 data points available for analysis.

Mean: The mean, or average, is computed by summing up all the values in the "age" variable and dividing by the count of data points. Here, the mean age is approximately 51.86, providing a measure of central tendency for the dataset.

Standard Deviation (std): The standard deviation gauges the spread or dispersion of the data, revealing how much individual data points deviate from the mean. For the "age" variable in this dataset, the standard deviation is approximately 16.97. A higher standard deviation signifies greater variability in ages.

Min: The minimum value in the "age" variable is 0, indicating that the youngest individual in the dataset has an age of 0.

25%: The 25th percentile, or the first quartile, denotes the value below which 25% of the data falls. In this dataset, 25% of individuals have an age of 40 or less.

50%: The 50th percentile, or the median, signifies the middle value when the data is sorted in ascending order. Here, 50% of individuals are 50 years old or younger.

75%: The 75th percentile, or the third quartile, represents the value below which 75% of the data falls. In this dataset, 75% of individuals have an age of 65 or less.
Max: The maximum value in the "age" variable is 85, indicating the age of the oldest individual in the dataset.These statistics provide a summary of the "age" variable in a dataset, offering insights into the distribution, central tendency, and spread of age values among the individuals in the dataset.

V. CONCLUSION

The examination of class imbalances within the HAM10000 dataset prompted the application of various mitigation techniques to enhance the fairness and efficacy of dermatological diagnostic models. This section presents the results and impact of implementing these strategies. Quantitative Metrics : The quantitative assessment of model performance reveals significant improvements resulting from class imbalance mitigation: Accuracy Enhancement : The overall accuracy of the model experienced a notable increase post-implementation of class imbalance mitigation techniques. This improvement underscores the model's heightened capability to accurately classify diverse skin disease instances. : recision, Recall, and F1-Score Improvements Precision, recall, and F1-score metrics were calculated across skin disease classes. The outcomes showcase a more equitable distribution of performance metrics, indicating a reduction in biases and improved diagnostic precision for underrepresented conditions. Visual Representations : Visualizations offer a comprehensive understanding of the impact of class imbalance mitigation techniques: Confusion Matrices: Confusion matrices visually depict the reduction in misclassifications and the improved true positive rates for minority classes. The enhanced model performance is evident in the minimized confusion among skin disease categories. ROC Curves :ROC curves demonstrate improved discrimination capabilities, with increased area under the curve (AUC) values. This enhancement signifies the model's ability to effectively distinguish between different skin disease classes, especially those with limited representation.

REFERENCES

[1] Ding, S., & Hu, C. (2023). Survey on the Convergence of Machine Learning and Blockchain. In K. Arai (Ed.), Intelligent Systems and Applications (Vol. 543, pp. 170–189). doi:10.1007/978-3-031-16078-3_10

[2] Meghla, T. I., Rahman, M. M., Biswas, A. A., Hossain, J. T., & Khatun, T. (2021, July). Supply Chain Management with Demand Forecasting of Covid-19 Vaccine using Blockchain and Machine Learning. 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 01–07. doi:10.1109/ICCCNT51525.2021.9580006

[3] Mondal, A., Virk, H., & Gupta, D. (n.d.). BEAS: Blockchain Enabled Asynchronous & Secure Federated Machine Learning.

[4] Leka, E., Hoxha, E., & Rexha, G. (2023, May). Security and Privacy Concerns Associated with the Internet of Things(IoT) and the Role of Adapting Blockchain and Machine Learning - A Systematic Literature Review. 2023 46th MIPRO ICT and Electronics Convention (MIPRO), 1690–1696. doi:10.23919/MIPRO57284.2023.10159869

[5] Miao, Y., Liu, Z., Li, H., Choo, K.-K. R., & Deng,R. H. (2022). Privacy-Preserving Byzantine-Robust Federated Learning via Blockchain Systems. IEEE Transactions on Information Forensics and Security, 17, 2848–2861.doi:10.1109/TIFS.2022.3196274

[6] Wang, Z., Duan, S., Wu, C., Lin, W., Zha, X., Han,P., & Liu, C. (2022, December). Generative Data Augmentation for Non-IID Problem in Decentralized Clinical Machine Learning. Retrieved from http://arxiv.org/abs/2212.01109

[7] Shalini, K. S., & Nithya, M. (2022). An integrated approach of block chain technology with machine learning and cloud computing for handling healthcare data. International Journal of Health Sciences, 7180–7190. doi:10.53730/ijhs.v6nS1.6545

[8] Guerra, E., Wilhelmi, F., Miozzo, M., & Dini, P. (2023). The Cost of Training Machine Learning Models Over Distributed Data Sources. IEEE Open Journal of the Communications Society, 4, 1111– 1126. doi:10.1109/OJCOMS.2023.3274394

[9] Liu, Y., Qu, Y., Xu, C., Hao, Z., & Gu, B. (2021).Blockchain-Enabled Asynchronous Federated Learning in Edge Computing. Sensors, 21(10), 3335. doi:10.3390/s21103335

[10] Ying, X., Liu, C., & Hu, D. (2023, July).GCFL: Blockchain-based Efficient Federated Learning for Heterogeneous Devices. 2023 IEEE Symposium on Computers and Communications (ISCC), 1033–1038.doi:10.1109/ISCC58397.2023.10218066

[11] Xu, M., Zou, Z., Cheng, Y., Hu, Q., Yu, D.,& Cheng, X. (2023). SPDL: A Blockchain-Enabled Secure and Privacy-Preserving Decentralized Learning System. IEEE Transactions on Computers, 72(2),548–558. doi:10.1109/TC.2022.3169436

[12] Li, J., Shao, Y., Ding, M., Ma, C., Wei, K., Han, Z., & Poor, H. V. (2020,December). Blockchain Assisted Decentralized Federated Learning (BLADE-FL) with Lazy Clients.Retrieved from http://arxiv.org/abs/2012.02044

[13] Bhat, S. A., Sofi, I. B., & Chi, C.-Y. (2020).Edge Computing and Its Convergence With Blockchain in 5G and Beyond: Security, Challenges, and Opportunities. IEEE Access, 8, 205340–205373. doi:10.1109/ACCESS.2020.3037108
[14] Singh, K. K., Balamurugan, B., Smieee, N. C., & Rawal Kshatriya, B. S. (2020). Machine Learning approaches for convergence of IoT and Blockchain. Open Computer Science, 10(1), 459–460. doi:10.1515/comp-2020-0207
[15] Bali, V., Bhatnagar, V., Bali, S., & Dahiya,N. (n.d.). Optimization and Convergence of Machine Learning Algorithms for Leveraging IoT, Blockchain, and Artificial Intelligence.
[16] AlDera, S. A., & Othman, M. T. B. (2022). A Model for Classification and Diagnosis of Skin Disease using Machine Learning and Image Processing Techniques. International Journal of Advanced ComputerScience and Applications, 13(5). doi:10.14569/IJACSA.2022.0130531
[17] Singh, D. D., Singh, A. K., & Srivastava, A. K. (2023). METHOD OF SKIN DISEASE DETECTION USING IMAGE PROCESSING AND MACHINE LEARNING. 11(7).
[18] R, S., & K, V. (2022, November). Skin Cancer prediction using Machine Learning. 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC), 1112–1115. doi:10.1109/IIHC55949.2022.10060544
[19] Jasti, V. D. P., Zamani, A. S., Arumugam, K., Naved, M., Pallathadka, H., Sammy, F., … Kaliyaperumal, K. (2022). Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis. Security and Communication Networks, 2022, 1–7. doi:10.1155/2022/1918379
[20] Kottath, A. V., & Shri Bharathi, S. (2022, September). Image Preprocessing Techniques in Skin Diseases Prediction using Deep Learning: A Review. 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), 1–6. doi:10.1109/ICIRCA54612.2022.9985547
[21] Abdulaal, M. J., Mehedi, I. M., Aljohani, A.J., Milyani, A. H., Mahmoud, M., Abusorrah, A. M., & Jannat, R. (2022). Separation of Different Blogs from Skin Disease Data using Artificial Intelligence. Computational Intelligence and Neuroscience, 2022, 1–8. doi:10.1155/2022/7538643
[22] K, M., Adepu, A., Nagandla, V. V. T., & Agarwal, S. (2023, June). A Machine Learning Based Melanoma Skin Cancer using Hybrid Texture Features. 2023 3rd International Conference on Intelligent Technologies (CONIT), 1–5. doi:10.1109/CONIT59222.2023.10205876
[23] Imad, M., Khan, Z. A., Hussain Bangash, S., Khan, I. U., Ahmad, S., & Ishtiaq, A. (2023, May). Comparative Analysis of Machine Learning Methods for Multi-Label Skin Cancer Classification. 2023 9th International Conference on Information Technology Trends (ITT), 181–186. doi:10.1109/ITT59889.2023.10184240
[24] Skin Disease Detection Using Machine Learning. (2023). International Journal of Food and Nutritional Sciences, 11(12). doi:10.48047/ijfans/v11/i12/171
[25] Marques, O., Tavares, J. M. R. S., & Celebi, M. E. (2019). Skin Cancer Recognition UsingConvolutional Neural Networks with Transfer Learning. Journal of Medical Systems, 43(9), 294.
[26] Kumar, A., Shukla, A., & Sharma, A. (2020). Enhanced Skin Lesion Classification with Edge Detection and Blockchain. Journal of Medical Imaging and Health Informatics, 10(7), 1624-1629.
[27] Li, Y., Zhang, X., Chen, S., & Wang, Y. (2021). Privacy-Preserving Medical Image Analysis Using Homomorphic Encryption and Blockchain. Journal of Medical Informatics, 15(4), 345-351.
[28] M. M. Sheeraz, M. A. I. Mozumder, M. O. Khan, M. U. Abid, M. -I. Joo and H. -C. Kim, "Blockchain System for Trustless Healthcare Data Sharing with Hyperledger Fabric in Action," 2023 25th International Conference on Advanced Communication Technology (ICACT), Pyeongchang, Korea, Republic of, 2023, pp. 437- 440, doi: 10.23919/ICACT56868.2023.10079423.
[29] Tanuwidjaja, H., Choi, R., Baek, S., & Kim,
[30] K. (01 2020). Privacy-Preserving Deep Learning on Machine Learning as a Service—a Comprehensive Survey. IEEE Access, 8, 167425–167447. doi:10.1109/ACCESS.2020.3023084
[31] G. A. Kusi, Q. Xia, C. N. A. Cobblah, J. Gao and H. Xia, "Training Machine Learning Models Through Preserved Decentralization," 2020 16th International Conference on Mobility, Sensing and Networking (MSN), Tokyo, Japan, 2020, pp. 465-472, doi: 10.1109/MSN50589.2020.00080.
[32] A. Saini, D. Wijaya, N. Kaur, Y. Xiang and L. Gao, "LSP: Lightweight Smart-Contract-Based Transaction Prioritization Scheme for Smart Healthcare," in IEEE Internet of Things Journal, vol. 9, no. 15, pp. 14005-14017, 1 Aug.1, 2022, doi: 10.1109/JIOT.2022.3145406.
[33] Z. A. E. Houda, A. S. Hafid, L. Khoukhi and B. Brik, "When Collaborative Federated Learning Meets Blockchain to Preserve Privacy in Healthcare," in IEEE Transactions on Network Science and Engineering, vol. 10, no. 5, pp. 2455-2465, 1 Sept.-Oct. 2023, doi: 10.1109/TNSE.2022.3211192.
[34] X. Fu, L. Bi, A. Kumar, M. Fulham and J. Kim, "Graph-Based Intercategory and Intermodality Network for Multilabel Classification and Melanoma Diagnosis of Skin Lesions in Dermoscopy and Clinical Images," in IEEE Transactions on Medical Imaging, vol. 41, no. 11, pp. 3266-3277, Nov. 2022, doi: 10.1109/TMI.2022.3181694.
[35] A. Magdy, H. Hussein, R. F. Abdel-Kader and K. A. E. Salam, "Performance Enhancement of Skin Cancer Classification Using Computer Vision," in IEEE Access, vol. 11, pp. 72120-72133, 2023, doi: 10.1109/ACCESS.2023.3294974.
[36] J. Wicaksana, Z. Yan, X. Yang, Y. Liu, L. Fan and K. -T. Cheng, "Customized Federated Learning for Multi-Source Decentralized Medical Image Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 11, pp. 5596-5607, Nov. 2022, doi: 10.1109/JBHI.2022.3198440.

[37] P. La Torraca, L. Ausiello, G. Zucchi, A. Farina and F. Pancaldi, "Identification of Soft Tissue-Mimicking Materials and Application in the Characterization of Sensors for Lung Sounds," in IEEE Sensors Journal, vol. 22, no. 1, pp. 1012-1019, 1 Jan.1, 2022, doi: 10.1109/JSEN.2021.3130546.
[38] X. Li, C. Desrosiers and X. Liu, "Deep Neural Forest for Out-of-Distribution Detection of Skin Lesion Images," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 1, pp. 157-165, Jan. 2023, doi: 10.1109/JBHI.2022.3171582.
[39] Fan, L., Zhang, F., Fan, H., & Zhang, C. (2019). Brief review of image denoising techniques. Visual Computing for Industry, Biomedicine, and Art, 2(1), 7. doi:10.1186/s42492-019-0016-7
[40] Wu, D., Lin, Z., Li, B., Ye, M., & Wang, W. (2017). Deep Supervised Hashing for Multi- Label and Large-Scale Image Retrieval. doi:10.1145/3078971.3078989
[41] Rodríguez-Torres, F., Martínez-Trinidad, J. F., & Carrasco-Ochoa, J. A. (2022). An Oversampling Method for Class Imbalance Problems on Large Datasets. Applied Sciences, 12(7). doi:10.3390/app12073424
[42] Alomar, K., Aysel, H. I., & Cai, X. (2023). Data Augmentation in Classification and Segmentation: A Survey and New Strategies. Journal of Imaging, 9(2). doi:10.3390/jimaging9020046
[43] Amarù, S., Marelli, D., Ciocca, G., & Schettini, R. (2023). DALib: A Curated Repository of Libraries for Data Augmentation in Computer Vision. Journal of Imaging, 9(10). doi:10.3390/jimaging9100232
[44] Komatsu, R., & Gonsalves, T. (2020). Comparing U-Net Based Models for Denoising Color Images. AI, 1(4), 465–486. doi:10.3390/ai1040029
[45] Lewy, D., & Mańdziuk, J. (2023). An overview of mixing augmentation methods and augmentation strategies. Artificial Intelligence Review, 56(3), 2111–2169. doi:10.1007/s10462- 022-10227-z
[46] Yutong Li, Ruoqing Zhu, Mike Yeh, Annie Qu. (2021). Dermoscopic Dataset for the "Dermoscopic Image Classification with Neural Style Transfer" Manuscript. IEEE Dataport. https://dx.doi.org/10.21227/10k4-3p31
[47] https://hex.tech/blog/data-preprocessing- steps-python/
[48] Wang, F., Vergara-Niedermayr, C., & Liu, P. (2014). Metadata based management and sharing of distributed biomedical data. International journal of metadata, semantics and ontologies, 9(1), 42–57. https://doi.org/10.1504/IJMSO.2014.059126
[49] https://www.labellerr.com/blog/guide-to- image-annotation-techniques-tools-and-best- practices/
[50] Nnamoko, N., & Korkontzelos, I. (2020). Efficient treatment of outliers and class imbalance for diabetes prediction. Artificial Intelligence in Medicine,104, 101815. doi:10.1016/j.artmed.2020.101815
[51] Hoens, T., & Chawla, N. (06 2013). Imbalanced Datasets: From Sampling to Classifiers. doi:10.1002/9781118646106.ch3
[52] Liu, J. (02 2022). Importance-SMOTE: a synthetic minority oversampling method for noisy imbalanced data. Soft Computing, 26. doi:10.1007/s00500-021-06532-4
[53] Fan, W., Si, Y., Yang, W., & Sun, M. (2022). Class-specific weighted broad learning system for imbalanced heartbeat classification. Information Sciences, 610, 525–548. doi:10.1016/j.ins.2022.07.074