ADVANCED JOURNAL OF ECONOMIC DATA ANALYTICS

Journal Website: https://journals.nwu.ac.za/index.php/ajeda/
DOI: https://doi.org/10.36386/9rjn6t31
ISSN: 3134-9587 (Print) | 3134-9595 (Online)
Vol. 1, Issue 1, 2026

Clinical features-based review of machine learning frameworks in predicting skin diseases

Olaniyan Ayodele Bukola[1], **Olawale-Success Olajumoke Oluwagbemisola**[2], **Ayobami Gabriel Ayeni**[3]

[1]*Department of Computer Science, Gateway (ICT) Polytechnic, Saapade, Ogun State, Nigeria.*
[2]*Department of Microbiology, New City University, Ayetoro, Ogun State, Nigeria.*
[3]*Department of Computer Science, Dominion University, Ibadan, Oyo State, Nigeria.*

Corresponding author's email: ayeni4great@yahoo.co.uk

Received: May 04, 2026 | **Accepted:** June 11, 2026 | **Published:** August 20, 2026

Abstract

Skin disease is a medical illness that is often overlooked or underestimated among chronic ailments and deadly disorders. Meanwhile, the geographic dispersion of inhabitants who constantly experience excessive temperature being occasioned by hot weather increases the risk of skin disease. Though data mining and machine learning techniques play a vital role in grouping labelled or unlabelled data based on shared characteristics or similar attributes into predefined classes, the size and quality of data are germane to classification. In this study, commonly applied techniques for recent machine learning frameworks in the detection of skin disease, and their empirical comparison, were explored with a functional and conceptual review. The existing methods and models were examined for their specific usage and applications. The hidden but relevant patterns in clinical data were examined for proper fitting of the classification model, with relative parameters and the target variable to improve the sensitivity and accuracy of each classifier. Machine learning embedded functions in prominent data analytic tools allow automation of computational processes, like model building and validation, to maintain the sensitivity of clinical datasets. However, most of the selected machine learning techniques from previous studies were faced with rarely available real-time medical data, which could not be directly used for clinical analysis because a larger proportion of the medical data requires pre-processing before feature selection.

Keywords: Clinical data, Diseases, ML framework, Prediction, Skin

1. Introduction

Skin disease is a common ailment but could be deadly like other ailments or disorders. People who are geographically dispersed in regions with excessive temperatures or occasionally induced hot weather are highly prone to the risk of skin diseases. Classification of skin disease is quite challenging; hence, identifying very relevant and optimal features for skin disease classification from huge clinical data is difficult (Ayeni et al., 2020).

Recent research in health informatics has shown the extensive application of computational intelligence or machine learning techniques in computer-assisted disease diagnosis based on a series of peculiar symptoms for various diseases and acute illnesses. Optimal use of healthcare applications in telemedicine during the global lockdown, particularly in the tense year 2020 due to the COVID-19 pandemic, brings about motivation for continuous development of computer-assisted diagnostic technologies or prognostic applications in healthcare and the medical domain (Usman et al., 2023).

At times, global warming and sudden changes in climate might be notable incidences which can cause skin diseases or some irritating skin rashes (Rangaswamy et al., 2024). The inward and outward thermal radiation of the sun is not far from the atmosphere when Earth revolves. Certain gases, radiation, carbon dioxide, wind patterns, methane, fossil fuels, and ocean currents are also attributed to factors influencing climate change and the warming of the planet (Chauhan, 2020). The suffering of people from different diseases has been alleviated through automated disease diagnosis and technological solutions such as pre-trained predictive models of supervised classification techniques (Sariyah et al., 2023; Meena et al., 2023).

Preventive measures and treatment of disease in a timely manner are very important in modern medicine, unlike the high number of preventable deaths recorded in the previous century due to the lack of accurate diagnosis of symptoms in patients, thereby threatening the patients' lives. Effective and timely diagnosis of any health-associated problem is of utmost importance, which necessitates standard diagnostic and clinical recovery for serious illness (Karwa et al., 2022). As much as classification algorithms and data mining techniques play a vital role in grouping labelled or unlabelled data based on shared characteristics or similar attributes into predefined classes, the size and quality of data are also germane (Zhang et al., 2024; Ibrahim et al., 2023).

The use of machine learning embedded functions in prominent data analytic tools helps in automating some computational processes, like model building and validation, to maintain the sensitivity of clinical datasets (Sheety et al., 2022; Ayeni, 2021). Limited availability of clinical data in real time is still not directly usable for clinical analysis because a larger proportion of the medical data requires extensive pre-processing before feature selection. Combining the efficacy of neural networks and fuzzy logic to optimize the functional perspective of a hybridized architecture brings an enhanced edge with machine learning techniques toward fundamental understanding and experimental analysis of clinical data (Ummapure et al., 2023).

As much as a large proportion of clinical data is needed to increase the training cycle for model improvement, pre-processing before the application of feature extraction techniques is germane. The pre-processing phase could involve filling missing values, converting fixed data to nominal values, and changing labels for medical data analysis (Hatem, 2022). Selecting the preferred techniques for recent machine learning frameworks in the detection of skin disease and their empirical comparison is the focus of this study and is further guided by two research questions as stated below:

1.1. Research questions

- **i.** What is the availability level of medical data for skin disease prediction and clinical analysis?
- **ii.** How can the hidden but relevant patterns in dermatological data be discovered for proper fitting?

2. Review of related literatures

Rangaswamy *et al.* (2024) provided a framework for classifying skin disease using a deep learning technique that incorporates a convolutional neural network (CNN). They codenamed two different networks as V3 and VGG16, which were trained and evaluated with seventeen thousand (17,000) images, as well as thirteen (13) classes. Image pre-processing was also done to validate the input data before experimenting with the dense layers of the CNN model, because understanding the epidemiological pattern of medical images, as shown in Figure 1, is very crucial. The accuracy of the classification algorithm was optimized; nevertheless, the study outcome provides insight into the automation of dermatological systems that could be adapted to other areas of medical informatics.

Figure 1: Lesion for Clinical Stages of Skin Disease (Raut et al., 2022)

An investigative and exploratory study by Sariyar (2023) juxtaposed the application of formal methods for data processing, preservation, and analysis for cause-and-effect relationships in clinical decision support with conjectural propositions of health informatics. The provision of appropriate principles, theories, and methods helps to substantiate the scientific notion of health or medical informatics; however, the uncertain perception of underlying philosophy is yet to align totally with interdisciplinary justification, unlike the sample exploration in the work of

Saran et al. (2023) conducted a descriptive survey which intended to examine the impact of health transformation policy on access equity to the usage of medical imaging service in Lorestan communities, western region of Iran. The annual growth of imaging devices was computed, while the equal distribution of medical images was determined for a target population close to one hundred thousand (100,000). Their results revealed a tremendous increase in the number of imaging devices through the overall allocation of imaging gadgets and the population ratio. However, access equity to basic health innovation is inadequate, despite the increase in these facilities. Stakeholders were advised to provide clinical facilities based on the specific health needs of different provinces.

Meena et al. (2023) proposed a novel method for predicting skin diseases using supervised classification techniques. Their work explored mining algorithms of K-nearest neighbour (KNN), support vector machine

(SVM), and random forest (RF) for the selection of principal attributes and removal of non-relevant features affecting the predictive performance of the method. Experimental analysis shows an intensified effort to increase the accuracy of the data mining method in feature selection with an ensemble of techniques to avoid model overfitting. However, comparative analysis with other predictive models or methods of similar systems was not provided, coupled with the fact that the work is a trial-and-error benchmark of raw skin data.

Usman et al. (2023) provided a mathematical approach to epidemic modelling with fear of viral infection. Numerical simulation of the proposed model leverages forward bifurcation to stabilize the free virus equilibrium in a dynamical study of COVID-19. Meanwhile, the risk level was not completely evaluated for the concentration of the virus toward the target population of infected humans. A multi-disease prediction system was developed using machine learning techniques, implemented by application prototyping to forecast transiting stages of various diseases since the data corpus for knowledge representation is based on the symptoms (Karwa et al., 2022). However, several localized diseases have specific features in various places, thereby making disease outbreak prediction somewhat difficult. Nevertheless, high fidelity, accuracy, and computational speed of a multi-disease predictive model are a great improvement.

Raut et al. (2022) applied machine learning algorithms to the classification of skin diseases using decision trees and support vector machines (SVM). Diagnostic accuracy was improved with the scalability of the medical decision support system. Meanwhile, expanding learning strategies for data pre-processing and hybridized techniques were identified as areas of further study. Ayeni (2021) also developed a clinical prognostic system for disease classification using fuzzy cluster means. The designed model for clinical prognosis used computational intelligence for handling ambiguity and imprecision in medical data, towards clinical classification and predictive projection of every ailment, but the predictive tendency of the proposed system was based on a mathematical model and/or a clustering algorithm for data mining. It produced an enhanced classification framework for a clinical prognostic system, which has given room for multi-class prediction, especially with large datasets in quantitative and developmental research.

Chauhan (2020) implemented a diagnostic system for disease prediction using machine learning. The study focused on the association of relevant attributes with the target for predicting several diseases to the advantage of human experts for smart decision support models. Medical information management and\or health informatics revolves around decision support model applications and diagnosis of various illnesses. However, delay in the diagnosis and treatment of serious ailments like diabetes and kidney failure could pose a high risk to one's life. Goswami et al. (2020) also conducted an image survey of skin disease classification from a descriptive and comparative perspective, which reveals how the classifier's efficiency depends largely on selected features. Some skin diseases have varying visual attributes, which make the selection of useful features from images more difficult. Adaptation of a pre-trained model with transfer learning brings augmentation to the existing approach.

3. Comparison review of selected frameworks

Advanced technologies of computing like deep learning and machine learning algorithms make it possible to predict skin diseases of various types. The accuracy of the results seems to improve when predictions and analysis are carried out, especially with data learning methods like Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Network (ANN), and Fuzzy Logic. However, sometimes the accuracy of the predictive model may appear biased or otherwise.

Data mining algorithms yield the functional framework, and the results of the developed solution system showed that Probabilistic Neural Networks (PNN) work better in precision for determining the critical stage in kidney disease, in comparison with three other algorithms. However, diagnostic and specificity error in treatment was only observed in Probabilistic Neural Networks (PNN). Skin disease diagnosis system using machine learning based on an ensemble of techniques helps to handle extracted features of principal components, while logistic regression was used for training and testing the clinical images. However, some difficulties were seen in the prediction process due to rough surface, irregular skin, scars, and other factors that are peculiar to image resolution and appearance.

Machine learning approaches to disease prediction in vulnerable patients in probabilistic nature of support vector machine (SVM) and k-nearest neighbour (KNN), were dependants of statistical learning for clinical features. Support vector machine (SVM) and fuzzy logic had been previously used together for accurate estimation of classification (Ibrahim et al., 2023). For instance, the yardstick for searching related instances in a clinical database or knowledge base in diagnostic decision support is similar to that of other systems. A comparative survey for analysing the existing algorithms for clustering and classification has been explored as well.

Essential yardsticks for controlling the inference engine from a graphical user interface include handling input/output elements, membership functions, rules, and monitoring the behaviour of specific rules for improving system performance with multi-dimensional inference or a multi-faceted adaptive system, as shown in Figures 2 to 5.

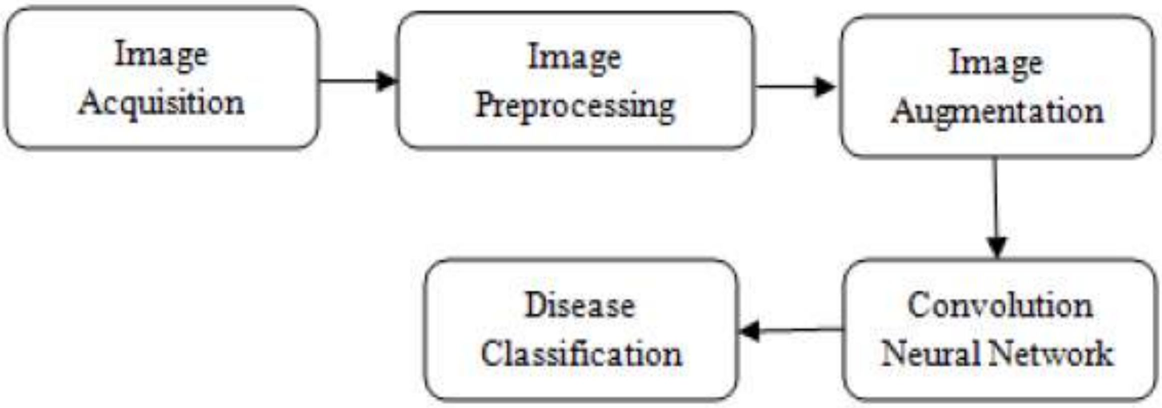

Figure 2: Workflow of Machine Learning classifier (Goswami et al., 2020)

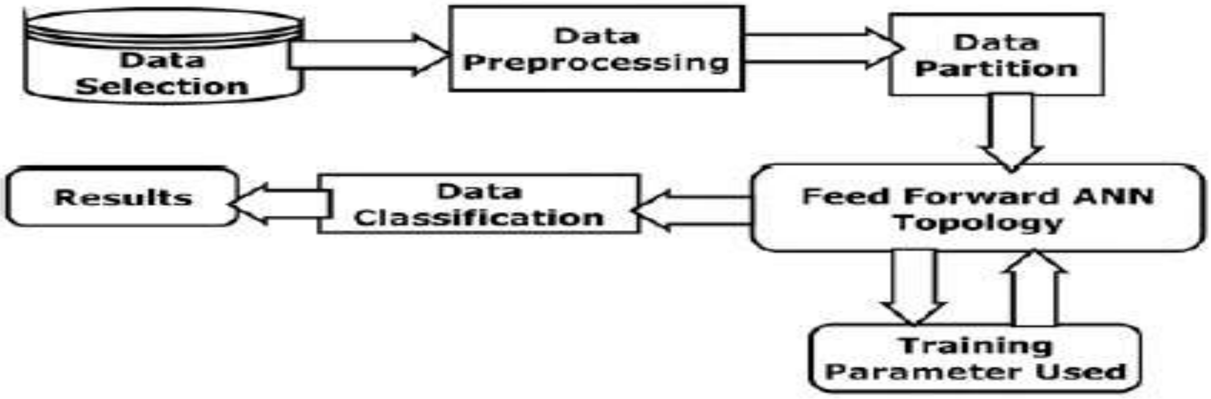

Figure 3: ANN–PSO Model for Feature Selection (Ahammed et al., 2022)

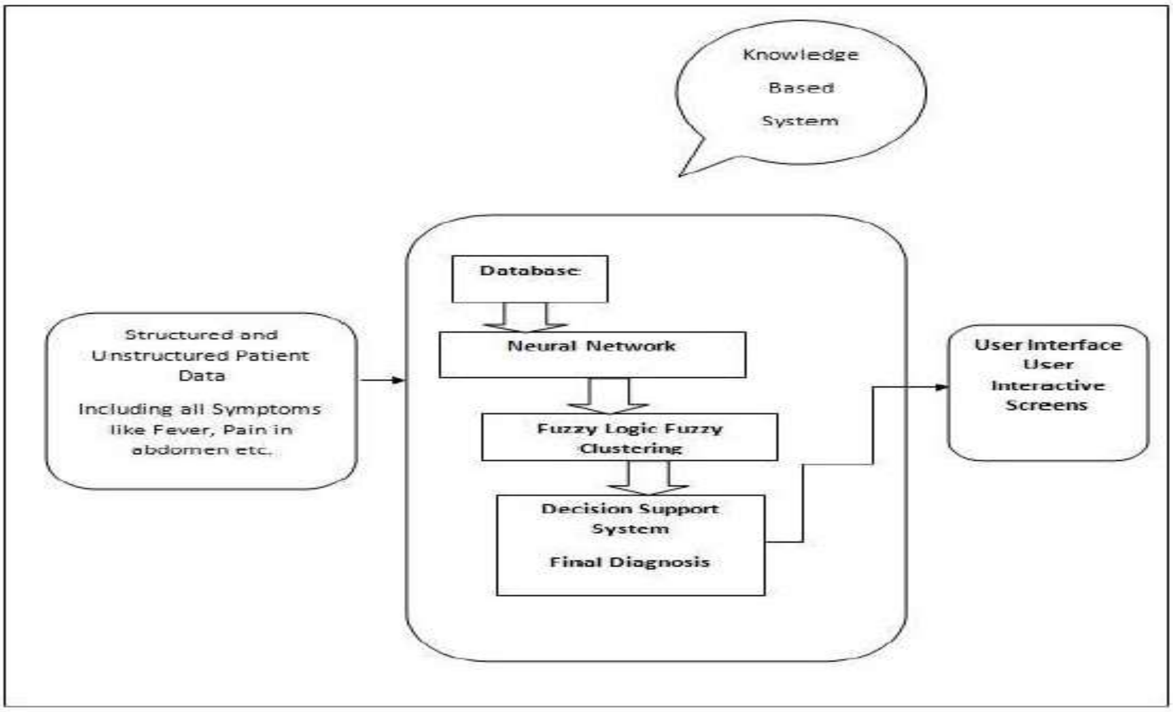

Figure 4: Neuro Fuzzy System for Diagnostic Support (Sayyad et al., 2022)

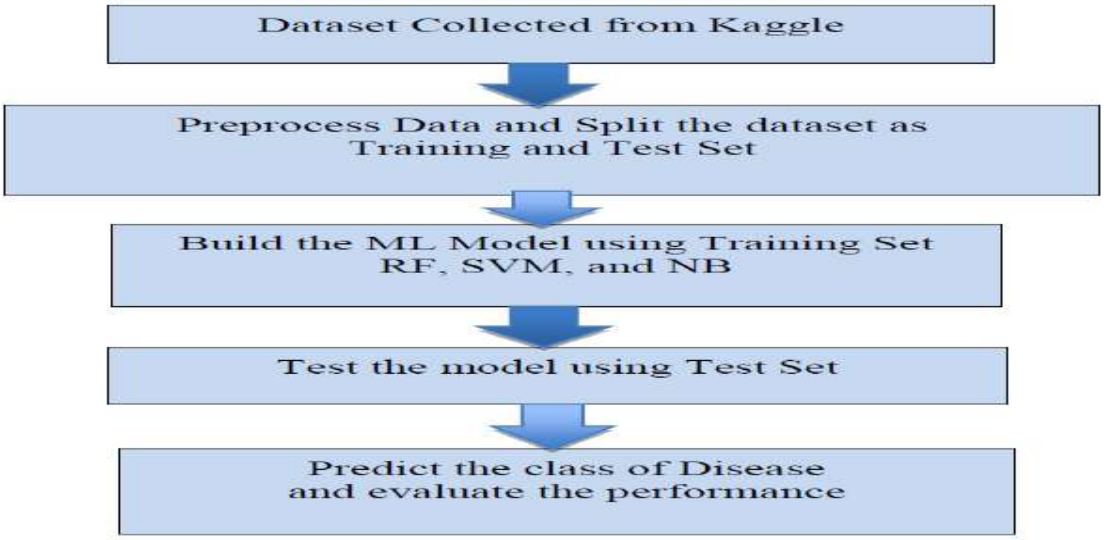

Figure 5: Multi-disease prediction workflow (Karwa et al., 2022)

Evaluating patients' data and clinical decisions for diagnostic support is very germane to the application of computational intelligence for classifying chronic diseases because classification methods help to prevent errors due to fatigue or lack of human expertise. The therapeutic dataset is subjected to dissection by quite a number of varying elements and restorative factual information. A computer-aided diagnosis blueprint makes it possible to gather the liver dataset from various hepatologists. Anomalies and imbalanced biomedical data are targeted with efficient implementation of a predictive model (Barin and Guraskin, 2022).

Feature selection methods are essential tools for model optimization and fitting in machine learning. An optimal paradigm for selecting an essential subset of features is needed to actualize overall performance when parameter tuning of algorithms is significant. Ensuring a high level of confidence because the classifier can capture the essential attributes and structure of the clinical domain. Its reasons from the rules so that it knows the system's internal processes and uses that knowledge to determine which state of the device matches actual observations. However, the main disadvantage of this approach is the computational inefficiency (Saifan and Jubair, 2022).

Dermatological characteristics and skin related features were used to define independent variables (input), as well as their membership function for fuzzy inference mechanism while skin status is being identified as the target variable (output) by reasoning, which does not only classify instance of skin data into normal and diseased but can explicitly create multi clusters and target groups by partitioning the clinical data to indicate the kind of skin disease being diagnosed. Validation can be performed by segmenting clinical data for training and testing, with operational peculiarities of different techniques as summarized in Table 1.

Applied Methods	**Strength (s)**	**Weakness (es)**
Deep learning technique	Accuracy of CNN and optimization of disease classifier	Need to automate dermatological system for adaptation
KNN, SVM and RF supervised technique	Selection of significant for model accuracy	Irrelative qualities and raw format of real time health data
Mathematical modeling approach	Numerical solution to infection stability	Risk level of infected human N/A
Empirical analysis	Emerging principles and theories within health informatics	Wrong perception of the underlying science philosophy
Descriptive survey	HTP influence on access usage of medical imaging	Lack of suitable data for analyzing the impact factor

Ensemble of machine learning techniques	High speed for disease prediction	No performance analysis/comparison
Decision tree C4.5 and Support vector machine	Scalability of decision support for diagnostic	Expansion of learning strategies after pre-processing
Feed forward neural network	Multi label clustering of clinical data	Limited parameter for model training
Neural network and fuzzy cluster means	Optimality in decision support	Lack of implementation guide
Computer aided disease diagnosis	Error prevention in system anomaly	Lack of scientific/ analytical experiment
Fuzzy cluster means	Multi class clinical data and prediction	Medical imaging not fully explored
Descriptive study and comparative survey	Adaptation of pre-trained model with transfer learning	Insight into classification dependency on data
Hybridized machine learning algorithms	Improved precision in diagnosis support	Lack of evaluation and comparison

Table 1: Summary of Functional Techniques for Selected Frameworks

4. Conclusion and Contribution to Knowledge

The computing approach reduces the challenges that conventional systems have, such as limited coverage in symptom and cause relationships and poor performance of the system when dealing with problems that lie on the periphery of its knowledge.

Consequently, classes of output or independent variables have been the common classification target of existing systems, such as normal and diseased, presence or absence, healthy and unhealthy, without estimating the intensity level. Hence, a mathematical model with the computational intelligence method of Neuro-fuzzy is more promising for increased efficiency in prognostic performance due to deep learning and scalability tendency, to inculcate the reasoning potential of machine learning techniques in a unique predictive framework.

A multi-layer inference system was incorporated in the adaptive Neuro-fuzzy technique for classifying dermatological data, though its prediction reliability for skin was not evaluated. Machine learning algorithms have been widely used in the past for disease classification in a comparative manner; the projection of the hybridized method is expected to produce perfect prediction with effectiveness.

Paucity of dermatological data in the health sector, especially data samples on skin diseases from clinical history of concentrated out-patients, is also a drawback; and when available contains too many irrelevant features or appears in unreadable form with hidden patterns, whereas classification efficacy of any algorithm depends majorly on the dataset and selected features.

Machine learning algorithms are heavily dependent on the qualitative data available to them and quantitative design for the learning and validation process to improve the model's behaviour, towards estimation of certainty in the target variable. Therefore, the given benchmark of raw medical data needs continuous enhancement of the classification model to adapt to contagious skin diseases.

Most of the selected machine learning techniques from previous studies were faced with rarely available real-time medical data, which could not be directly used for clinical analysis because a larger proportion of the medical data requires pre-processing before feature selection. The use of multiple efficient data mining techniques is shown to reveal and extract hidden information from clinical and laboratory patient data, which can be helpful to assist physicians in maximizing accuracy for identification of disease severity stage.

Therefore, data mining and machine learning algorithms play a major role in extracting hidden data from the large patient medical and clinical datasets that physicians frequently collect from patients to obtain insights about the diagnostic information and to implement treatment plans.

References

1. Ahammed, M., Al Mamun, M. and Uddin, M.S., 2022. A machine learning approach for skin disease detection and classification using image segmentation. *Healthcare Analytics*, *2*, p.100122.

2. Ayeni, A.G., 2025. Prognostic system for clinical classification of liver disease using fuzzy cluster means. *Journal of Statistical Sciences and Computational Intelligence Volume*, *1*(3), pp.187-197.
3. Barın, S. and Güraksın, G.E., 2022. An automatic skin lesion segmentation system with hybrid FCN-ResAlexNet. *Engineering Science and Technology, an International Journal*, *34*, p.101174.
4. Gandhi, K., Mittal, M., Gupta, N. and Dhall, S., 2020. Disease prediction using machine learning.
5. Goswami, T., Dabhi, V.K. and Prajapati, H.B., 2020, March. Skin disease classification from image-a survey. In *2020 6th international conference on advanced computing and communication systems (ICACCS)* (pp. 599-605). IEEE.
6. Hatem, M.Q., 2022. Skin lesion classification system using a K-nearest neighbor algorithm. *Visual Computing for Industry, Biomedicine, and Art*, *5*(1), p.7.
7. Ibrahim, A.M., Elbasheir, M., Badawi, S., Mohammed, A. and Alalmin, A.F.M., 2023. Skin cancer classification using transfer learning by VGG16 architecture (case study on Kaggle dataset). *Journal of Intelligent Learning Systems and Applications*, *15*(3), pp.67-75.
8. Karwa, H., Gupta, P., Agrawal, R., Virdi, G.S., Kumar, A. and Jain, S., 2022. Multi-disease prediction with machine learning. *International Journal of Health Sciences*, *6*(S2), pp.9477-9483.
9. Rangaswamy, S., Tantry, S.S. and Lal, T.S., 2025. Skin disease classification using deep learning. *National Academy Science Letters*, *48*(5), pp.585-588.
10. Raut, R., Borole, Y., Patil, S., Khan, V.N. and Takale, D.G., 2022. Skin disease classification using machine learning algorithms. *NeuroQuantology*, *20*(10), pp.9624-9629.
11. Saifan, R. and Jubair, F., 2022. Six skin diseases classification using deep convolutional neural network. *International Journal of Electrical and Computer Engineering*, *12*(3), p.3072.
12. Saran, M., Teli, B.D., Rezapour, A., Motlagh, S.N., Behzadifar, M., Haghighatfard, P., Bragazzi, N.L. and Behzadifar, M., 2023. The impact of the Iranian health transformation plan policy on equitable access to medical imaging services in West Iran. *BMC Research Notes*, *16*(1), p.350.
13. Sariyar, M., 2023. Is Medical Informatics a Scientific Discipline or Just Applied Computer Science?. In *Healthcare Transformation with Informatics and Artificial Intelligence* (pp. 385-389). 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP: SAGE Publications.
14. Sayyad, J., Patil, P. and Gurav, S., 2023. Skin disease detection using VGG16 and InceptionV3. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(1s), pp.148-155.
15. Shetty, B., Fernandes, R., Rodrigues, A.P., Chengoden, R., Bhattacharya, S. and Lakshmanna, K., 2022. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific reports*, *12*(1), p.18134.
16. Ummapure, S.B., Tilekar, R. and Mallappa, S., 2023. Skin cancer classification using VGG-16 and Googlenet CNN models. *International Journal of Computer Applications*, *184*(42), pp.5-9.
17. Usman, I.G., Ibrahim, M.O., Isah, B.Y., Lawal, N. and Akinyemi, S.T., 2023. Application of non-standard finite difference method on COVID-19 mathematical model with fear of infection. *Fudma Journal of Sciences*, *7*(4), pp.357-368.
18. Veni, N.K., Deepapriya, B., Vardhini, P.H., Kalyani, B. and Sharmila, L., 2022. A Novel Method for Prediction of Skin Diseases Using Supervised Classification Techniques. *Skin*, *20*, p.21.
19. Yekinni, A., 2020. Electrocardiogram (ECG) Signal and Interference Filtering for Clinical Diagnostic Support.
20. Zhang, X., Wang, S., Liu, J. and Tao, C., 2018. Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. *BMC medical informatics and decision making*, *18*(Suppl 2), p.59.